



**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE ESTUDOS DA LINGUAGEM**

LEONARDO COUTO FRANCO DE OLIVEIRA

**LÉXICO, ALOFONIA E PERCEPÇÃO DE FALA NA
FONOLOGIA ARTICULATÓRIA**

**CAMPINAS,
2003**

LEONARDO COUTO FRANCO DE OLIVEIRA

**LÉXICO, ALOFONIA E PERCEPÇÃO DE FALA NA
FONOLOGIA ARTICULATÓRIA**

Dissertação de Mestrado apresentada ao
Instituto de Estudos da Linguagem da
Universidade Estadual de Campinas para
obtenção do título de Mestre em
Linguística.

Orientadora: Profa. Dra. Eleonora Cavalcante Albano

**Este exemplar corresponde à versão
final da Dissertação defendida pelo aluno
Leonardo Couto Franco de Oliveira e orientada
pela Profa. Dra. Eleonora Cavalcante Albano.**

**CAMPINAS,
2003**

Agência(s) de fomento e nº(s) de processo(s): FAPESP, 01/00137-9

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Estudos da Linguagem
Crisllene Queiroz Custódio - CRB 8/8624

OL4L Oliveira, Leonardo Couto Franco de, 1977-
Léxico, alofonia e percepção de fala na fonologia articulatória / Leonardo Couto Franco de Oliveira. – Campinas, SP : [s.n.], 2003.

Orientador: Eleonora Cavalcante Albano.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Estudos da Linguagem.

1. Fonologia. 2. Percepção da fala. 3. Alofones. 4. Lexicologia. I. Albano, Eleonora Cavalcante, 1950-. II. Universidade Estadual de Campinas. Instituto de Estudos da Linguagem. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Lexicon, allophony and speech perception in articulatory phonology

Palavras-chave em inglês:

Phonology

Speech perception

Allophones

Lexicology

Área de concentração: Linguística

Titulação: Mestre em Linguística

Banca examinadora:

Eleonora Cavalcante Alvabo

Plínio Almeida Barbosa

Hani Camille Yehia

Data de defesa: 25-07-2003

Programa de Pós-Graduação: Linguística



BANCA EXAMINADORA:

Eleonora Calvacante Albano

Hani Camille Yehia

Plínio Almeida Barbosa

**IEL/UNICAMP
2003**

Ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós Graduação do IEL.

Pluralitas non est ponenda sine necessitate
(Não se deve recorrer à pluralidade sem necessidade)
– William de Occam

Agradecimentos

Como um trabalho essencialmente experimental e inserido no trabalho de um grupo, este trabalho se beneficiou das idéias, auxílios, e amizade de várias pessoas. São elas:

A professora Eleonora Cavalcante Albano, orientadora e mestra que sempre guiou com muita liberdade e precisão minhas investigações. Sua dedicação e rigor me estimularam a buscar sempre um pouco mais.

Os professores Plínio A. Barbosa, Ana Luiza G. P. Navas e Edson Françoze, excelentes professores e críticos.

O professor Fábio Violaro, que colaborou decisivamente para que atingíssemos uma qualidade superior na realização dos experimentos.

Gustavo Conde, locutor dos experimentos.

Os sujeitos que participaram voluntariamente dos experimentos.

Os amigos do LAFAPE, discentes e alguns já docentes: Beatriz Raposo de Medeiros, Adelaide Silva, Patrícia Aquino, Aglael Gama Rossi, Luciana Ruiz, Pablo Arantes, Laudino Roces, Renato Basso, Orlando Coelho, Maria Luiza Cunha Lima, Martina Marana, Ricardo Garcia, Jussara Vieira, Alexsandro Meireles e Leandro Silveira.

Os amigos de Campinas e *elsewhere*: Gregório, Anaia, Gabriel, Daniela, Max, Cristina, Alexandre, Pablo, Pedro, Cristina, Victória, Marcelo, Elaine, Ricardo Munduruca, Wellington, Geraldo Noel, Alexandre Caroli, Guilherme Nicésio, Luís Fernando, Rita, Marina, Débora, Suely, Pedro, Cláudia e Valter.

E além das pessoas, agradeço também à CAPES, FAEP e FAPESP (processo 01/00137-9) pelo financiamento da pesquisa e de meus estudos. Este trabalho também se beneficiou do financiamento ao Projeto Temático FAPESP “Integrando parâmetros contínuos e discretos em modelos do conhecimento fônico lexical” (processo 01/00136-2).

Resumo

A variante posicional de um fonema é chamada de alofonia. Geralmente considerada como processo fonético de aplicação automática e não consciente para os falantes da língua. No entanto, resultados preliminares da literatura indicam que variantes alofônicas podem ser percebidas diferentemente em palavras e não palavras. Seria, portanto, possível, que um fenômeno fonético como a alofonia poderia ser influenciado pelo status lexical? Para testar essa hipótese, realizaram-se dois experimentos perceptuais com a alofonia de /a/ no português do Brasil em posição acentuada e pós-acentuada. Foram coletados dados de tempo de reação e taxa de acerto em paradigma de discriminação ABX. Os estímulos foram com vogais naturais e manipuladas para os contextos acentuais. Os contextos de vizinhança lexical, frequência de ocorrência e probabilidade fonotática foram controlados no segundo experimento. No total, foram coletados dados de 18 falantes nativos do português do Brasil. Os resultados do experimento sugerem que há mais facilidade na discriminação dos alofones a depender da lexicalidade e do contexto acentual, na direção da hipótese inicial. Os resultados são discutidos à luz das Teorias da Fonologia Articulatória, Fonologia e da Teoria de Exemplares.

Abstract

The positional variant of a phoneme is known as allophony. It is usually considered a phonetic process of automatic and unconscious execution by speakers of a language. However, preliminary results indicate that allophonic variants can be perceived differently in words and non-words. Would it, then, be possible, that a phonetic phenomenon, such as allophony, could be influenced by lexical status? In order to test this hypothesis, two perceptual experiments with vowel /a/ allophony in stressed and post-stressed position in Brazilian Portuguese were performed. The collected data consisted of reaction time and correct response rate in an ABX discrimination paradigm. The stimuli consisted of naturally produced and modified vowels to match stress context. Lexical neighborhood, frequency of occurrence e phonotactical probability were also controlled for the second experiment. In total, data from 18 native speakers of Brazilian Portuguese subjects were collected. The results suggest it is easier to distinguish allophones depending on lexical status and stress context, as per the initial hypothesis. Results are discussed under the viewpoint of the theories of Articulatory Phonology and Exemplar Theory.

Lista de Figuras

2.1	Espectrograma de [pɐ'ga] concatenado, experimento 1.	24
2.2	Espectrograma de [pa'ga] sem concatenação, experimento 1.	24
2.3	Histograma do tempo de reação do experimento 1 com dados para todos os sujeitos e resultados dos teste de normalidade <i>Kolmogorov-Smirnov</i> e <i>Lilliefors</i>	28
2.4	Gráfico de probabilidade normal do experimento 1 com dados para todos os sujeitos.	28
2.5	Histograma do tempo de reação do experimento 1 com dados normalizados para todos os sujeitos.	30
2.6	Gráfico de probabilidade normal do experimento 1 com dados normalizados para todos os sujeitos.	31
2.7	Correlação entre o tempo de reação normalizado e o tempo de reação bruto para todos os sujeitos e resultados dos teste de normalidade <i>Kolmogorov-Smirnov</i> e <i>Lilliefors</i>	32
2.8	Quantidade de acertos/erros para as tríades classificadas por quantidade de formas incorretas, experimento 1.	34
2.9	Frequência de ocorrência dos pares /paga paga(r)/ e /daba daba(r)/, medida de vizinhança fonológica e número absoluto de vizinhos fonológicos (ver capítulo 3).	36
3.1	Correlação entre $\log_{10} f$ e média dos tempos de reação para 180 palavras do PB numa tarefa de decisão lexical a partir de leitura.	38
3.2	Lei de <i>Zipf</i> para as 20 mil palavras mais freqüentes do CETENFolha.	39
3.3	Correlação entre SFV e o número absoluto de vizinhos para 24578 palavras de até 14 fones do PB.	43
3.4	Superfície de contorno (<i>spline</i>) para frequência de ocorrência e SFV por tempo de reação	45
4.1	Distribuição de ['pagɐ] e ['vape] na distribuição normal teórica de SFV para quatro segmentos.	50
4.2	Distribuição de [pa'ga] e [va'pa] na distribuição normal teórica de SFV para cinco segmentos.	51
4.3	Distribuição de ['pagɐ] e ['vape] na distribuição normal teórica de Fonotaxe probabilística para quatro segmentos.	51
4.4	Distribuição de [pa'ga] e [va'pa] na distribuição normal teórica de Fonotaxe probabilística para cinco segmentos.	52
4.5	Quantidade de acertos/erros para as tríades classificadas por quantidade de formas incorretas, experimento 2.	54

Lista de Tabelas

2.1	Padrão de erro médio por bloco para os blocos e sujeitos, experimento 1. .	27
2.2	Valores de média, mediana, desvio padrão, faixa de variação (em ms), assimetria e curtose para os tempos de reação brutos e normalizados, experimento 1.	30
2.3	Padrão de acertos/erros para os fatores independentes binários, experimento piloto.	34
4.1	Valores de média, mediana, desvio padrão, faixa de variação (em ms), assimetria e curtose para os tempos de reação brutos e normalizados, experimento 2.	53
4.2	Padrão de erros para os blocos e sujeitos, experimento 2.	53
4.3	Padrão de acertos/erros para os fatores independentes binários, experimento 2.	54

Sumário

1	Introdução	13
1.1	Alofonia e representação lexical	14
1.2	Alofonia e Fonologia Articulatória	15
1.3	O experimento de Whalen, Best & Irwin 1997	17
1.4	Alofonia de [a, ɐ]	19
2	Experimento 1: Efeitos de variáveis lexicais na percepção da alofonia	21
2.1	Metodologia	22
2.1.1	<i>Corpus</i> utilizado	22
2.1.2	Sujeitos	23
2.1.3	Estímulos	23
2.1.4	Variáveis	25
2.1.5	Procedimentos	26
2.2	Análise Estatística	26
2.2.1	Análise Estatística do tempo de reação	27
2.2.2	Análise Estatística dos acertos/erros	32
2.3	Resultados	33
2.3.1	Resultados de tempo de reação	33
2.3.2	Resultados de padrão de acertos/erros	33
2.4	Discussão	35
3	Frequência de ocorrência, vizinhança lexical e fonotaxe	37
3.1	Frequência de ocorrência das palavras	37
3.2	Vizinhança fonológica	40
3.3	Fonotaxe probabilística	44
4	Experimento 2: extensão e aperfeiçoamento	47
4.1	Metodologia	47
4.1.1	<i>Corpora</i> utilizados	47
4.1.2	Sujeitos	48
4.1.3	Estímulos	49
4.2	Controles	50
4.2.1	Controle de vizinhança fonológica	50
4.2.2	Controle de fonotaxe probabilística	50
4.2.3	Variáveis	50
4.2.4	Procedimentos	52
4.3	Análise Estatística	52
4.4	Resultados	53
4.4.1	Resultados de tempo de reação	53

4.4.2	Resultados de padrão de acertos/erros	53
4.5	Discussão	55
5	Discussão	56
5.1	Discussão dos resultados	56
5.2	Fonologia Articulatória e teoria de exemplares	58
5.3	Lexicalidade: quais as fronteiras?	59
	Referências bibliográficas	61
	Apêndices	67
	Ficha de participação	67
	Instruções para os experimentos	67
	Locuções gravadas para o experimento 1	68
	Locuções gravadas para o experimento 2	68

Capítulo 1

Introdução

O objetivo deste capítulo introdutório é apresentar o fenômeno principal a ser discutido, a alofonia. Entendida basicamente como a variante contextual de uma unidade fonológica do tamanho do fonema (TRUBETZKOY, 1969), a alofonia é um conceito utilizado para descrever variações fônicas contínuas, tradicionalmente tratadas pela Fonética, ou discretas, tradicionalmente tratadas pela Fonologia. Argumentaremos que a alofonia tratada através de regras é uma abordagem que supõe um léxico mental econômico como o proposto pela Fonologia Gerativa (seção 1.1) e que a Fonologia Articulatória (doravante FAR) pode explicar alofonias sem postular regras simbólicas atemporais necessitando, porém, de um léxico mais explícito que o da Fonologia Gerativa. Essa explicitude necessária tem como consequência esvaziar o valor epistemológico do princípio de economia (seções 1.2 e 1.4). Na seção 1.2, introduziremos a discussão sobre a especificação da alofonia na FAR, que é ponto central do trabalho. Como uma possibilidade de investigação empírica do papel da alofonia na representação lexical, apresentaremos o estudo de (WHALEN, D. H.; BEST; IRWIN, 1997) (seção 1.3) sobre a percepção da alofonia de [p, p^h] no inglês americano, estudo no qual nos baseamos para desenvolver os dois experimentos apresentados neste trabalho. Na seção final apresentaremos a alofonia de [a, ɐ] no português do Brasil (doravante PB), alofonia que foi a escolhida para investigarmos em que medida a representação lexical poderia incluir alofones em uma teoria dinâmica como a FAR (seção 1.4).

1.1 Alofonia e representação lexical

A abordagem gerativa da Fonologia, que tem como obra de referência *The Sound Pattern of English* de Chomsky & Halle (1968), substitui o fonema por feixes de traços articulatorios e coloca a alofonia como um fenômeno governado por regras de derivação simbólica. Em *Sound Pattern of English*, as mesmas regras que derivam processos fonológicos como acentuação lexical, assimilação e mudança fonológica morfológicamente motivada, derivam feixes de traços (fonemas) em alofones. Tal tratamento indica que as variantes dos fonemas/feixes de traços não precisam estar lexicalmente especificadas, i.e., definidas subjacentemente. Regras dependentes de contexto seriam suficientes para prever as possibilidades de ocorrência de cada fonema ou feixe de traços em qualquer palavra do léxico. A posição de *Sound Pattern of English* continua inalterada em essência nas abordagens fonológicas gerativas posteriores:

“Such variants (*allophones*) of the same underlying sound (*phoneme*) are the product of systematic rules that modify the segment depending on the context in which it finds itself. [...] We are in general unconscious of these rules.” ((KENSTOWICZ, 1994, p. 66), grifo do autor)

A abordagem gerativa para o fenômeno da alofonia, apesar de econômica e simples possui, entretanto, ao menos duas premissas que não são claramente confirmadas empiricamente.

A primeira é a premissa de que regras se aplicam em todo o léxico e de forma puramente estrutural, i.e., tendo em conta apenas contextos fonológicos, morfológicos, sintáticos. Evidência contrária a essa premissa é apresentada por Luce e colegas (1999), que investigaram a produção e percepção de [t, d] e [r] como alofone de ambos no inglês americano. Os autores encontraram resultados que sugerem que o alofone [r] pode ser mapeado numa categoria distinta de [t, d] no lado da produção e não parece possuir uma representação independente no lado perceptual. Sugerem ainda que modelos de reconhecimento de palavras devam incorporar tanto aspectos abstratos, como na Fonologia gerativa, quanto aspectos de detalhe fonético fino. Veremos (capítulo 3) que variáveis como frequência de ocorrência, vizinhança fonológica e fonotaxe também influenciam a percepção de fala, sugerindo um léxico menos uniforme.

A segunda é a premissa de que as alofonias se apresentam como categóricas e discretas. Neste caso, a literatura fonética provê bastante evidência de que diversos casos de alofonia não podem ser tratados como um fenômeno do tipo tudo ou nada. Como exemplos bem documentados temos os alofones *clear* e *dark* do /l/ do inglês americano (SPROAT; FUJIMURA, 1993), róticos em PB (SILVA, 2002) e /l/ intervocálico em inglês americano (HUFFMAN, 1997).

Um possível desafio para uma teoria fonológica que se pretendesse alternativa às Fonologias gerativas seria então, o de, além de levar em conta o contexto estrutural de ocorrência, incorporar variáveis como frequência de ocorrência e prover a possibilidade de alofonias contínuas. Admitindo, é claro, que as características não-estruturais e de gradência sejam relevantes para a descrição e explicação fonológicas, abolindo a separação clássica entre Fonética (como implementação) e Fonologia (como sistema). As teorias gerativas, incluindo a Teoria da Otimalidade, restringem-se ao contexto estrutural, deixando para a implementação fonética os outros aspectos mencionados. O que questionaremos mais adiante (capítulo 3) é até que ponto ganhamos em poder explicativo ao incluir variáveis classificadas tradicionalmente como fonéticas e de desempenho na especificação lexical e fonológica.

1.2 Alofonia e Fonologia Articulatória

Uma alternativa ao tratamento dado pelas Fonologias gerativas ao fenômeno da alofonia e dos processos fonológicos em geral é o modelo de descrição e explicação fonológica proposto pela FAR¹ (BROWMAN; GOLDSTEIN, 1992). Na FAR, as unidades fonológicas mínimas são os gestos articulatórios, i.e., padrões motores abstratos dotados de tempo intrínseco que permitem especificar tanto o detalhe fonético quanto padrões globais de articulação. As entradas lexicais são definidas não mais segmentalmente ou a partir de traços articulatórios, mas a partir das chamadas pautas gestuais, que são conjuntos de gestos com coordenação temporal e magnitude espacial especificadas lexicalmente. Assim, o que determina, tanto no nível abstrato quanto no nível de execução motora como as formas são pronunciadas é a presença ou ausência dos gestos articulatórios dotados de duração intrínseca.

¹Uma introdução à teoria, e que também inclui propostas de extensão nos domínios prosódico e morfológico, encontra-se em Albano (2001).

Além da especificação da pauta gestual, a FAR utiliza dois princípios básicos para dar conta dos processos fonológicos em geral:

Princípio 1 *Gestos articulatórios podem variar seu grau de sobreposição, ou fase relativa.*

Princípio 2 *Gestos articulatórios podem variar sua magnitude temporal e espacial.*

Em relação ao princípio 1, uma das grandes críticas à FAR é a de que a fase relativa pode ter especificações potencialmente infinitas, o que não é desejável numa especificação que se pretende abstrata. Albano (2001) propõe que no nível da especificação gestual da entrada lexical, os gestos se alinham com as bordas dos gestos adjacentes, sem necessariamente possuir uma relação de fase rígida. A proposta simplifica e restringe o leque de combinações de sobreposição gestual, deixando algum espaço para variações de pronúncia de falantes, de itens lexicais com pronúncia marcada, etc.

Propõe-se que ambos os princípios substituam as operações de inserção, apagamento, assimilação, entre outras, propostas pelas Fonologias gerativas, numa teoria sujeita a princípios da Teoria dos Sistemas Dinâmicos (FOWLER, 1980; PORT; GELDER, 1995; KELSO, 1995), o que torna indesejáveis soluções em que haja derivação de processos por regras simbólicas atemporais.

Vejamos alguns exemplos de como os dois princípios básicos da FAR podem explicar processos fonológicos. No PB, temos em alguns dialetos verbos no gerúndio como “cantando” pronunciados como [kẽⁿˈtẽnũ]. Numa abordagem gerativa, o fenômeno poderia ser explicado por uma regra de apagamento do segmento [d] ou assimilação da consoante [d] do traço de nasalidade do arquifonema /N/ (MATTOSO CÂMARA JR., 1970). Já, numa abordagem como a FAR, um deslizamento e conseqüente sobreposição do gesto nasal com o gesto consonantal de ponta de língua para o [d] resultaria no mesmo efeito acústico, sem recorrer ao apagamento de segmentos ou à assimilação de nasalidade pela consoante seguinte. Ainda que possíveis, as duas últimas alternativas são consideradas processos fonológicos marcados pela Fonologia Gerativa.

Outro exemplo, desta vez utilizando o princípio de redução de magnitude, é o tratamento da redução de [t/d] para [ɾ] (*flapping*) no inglês americano. Intuitivamente, o que existe é um enfraquecimento com conseqüente redução da duração de [t/d], resultando em [ɾ]. Em abordagens gerativas, o processo pode ser explicado basicamente como uma

transformação de [t/d] por [r]. Vaux (2002) enumera mais de 30 propostas baseadas em traços articulatórios, nas quais a principal variação é o contexto de aplicação da regra. Já a explicação via FAR (BROWMAN; GOLDSTEIN, 1992) postula que há redução na magnitude do gesto de ponta de língua. A redução na magnitude seria tanto de deslocamento espacial, quanto na dimensão temporal. A vantagem é clara: sem a necessidade de transformações, que podem ser vistas também como substituições, a redução da “intensidade” do gesto e sua menor duração são explicadas por um mesmo princípio geral. No caso da duração, as análises gerativas nem chegam a tentar explicá-la, já que tais fenômenos tendem a ser vistos como resultantes de um nível prosódico separado do segmental.

Não parece custoso defender a FAR como alternativa para explicar tais tipos de alofonia, principalmente quando se leva em conta que a teoria tem em um de seus pontos fortes prever fenômenos de fala que se mostram gradientes e parecem ser pervasivos também nas alofonias (BROWMAN; GOLDSTEIN, 1992). Algumas alofonias, no entanto, não podem ser explicadas pelos princípios enumerados (1 e 2). Casos como o da variação de /r/ inicial em dialetos do Sul do Brasil, realizados como [r] ou [ɣ] pelo mesmo falante (SILVA, 2002), na mesma palavra, não são passíveis de explicação pelos princípios expostos, já que envolvem diferentes gestos articulatórios. Silva (2002), propõe que a diferença entre as variantes seja também uma diferença de especificação de gestos articulatórios no léxico: [ɣ] teria um gesto dorsal crítico não presente em [r].

1.3 O experimento de Whalen, Best & Irwin 1997

O estudo de Whalen e colegas (doravante WBI) parte da constatação de que as teorias fonológicas tradicionais não fazem previsões explícitas sobre a percepção de alofones, mas que, por não assinalarem distinções lexicais e, portanto, pertencerem a uma única categoria, a previsão seria de que os falantes teriam dificuldade de perceber os alofones como pertencendo a categorias diferentes. A expectativa dos autores é, ao contrário, a de que alofones em contexto não-apropriado sejam percebidos como exemplares piores do que alofones em sua posição apropriada. Foram realizados, ao todo, cinco experimentos para investigar se as diferenças alofônicas contextuais de /p/ em logatomas e palavras reais estariam disponíveis aos falantes tanto na produção quanto na percepção de fala. A alofonia de /p/ em inglês é condicionada principalmente pelo acento lexical e pela

posição inicial de palavra. Em posição acentuada e em inicial, realiza-se [p^h], em posição não-acentuada, [p].

No primeiro experimento, investigou-se a discriminabilidade e classificação de alofones em posição correta e incorreta. Os estímulos foram dissílabos com posição acentual variada e foi utilizada fala natural produzida por um foneticista com os alofones em posição correta e incorreta. Para o experimento de discriminação, empregou-se o paradigma *AXB*² e, para o teste de classificação, pediu-se que os sujeitos classifikassem os alofones numa escala que varia de um membro ruim da categoria (grau 1) até um bom membro da categoria (grau 7).

No segundo experimento, modificou-se apenas o teste de classificação, pedindo-se que os participantes classifikassem os estímulos por grau de sotaque estrangeiro.

O terceiro experimento investigou a produção dos alofones em contexto apropriado e inapropriado, pedindo-se que os sujeitos simplesmente repetissem o item numa condição, i.e., pronunciassem o mais rapidamente possível, e o imitassem na outra, i.e., repetissem com precisão fonética na pronúncia.

No quarto experimento, investigou-se a percepção dos alofones em seis palavras com estrutura silábica semelhante aos logatomas, também com alofones em posição correta e incorreta, utilizando apenas o teste de classificação com o mesmo critério do segundo experimento.

Finalmente, o último experimento testou a imitabilidade das palavras utilizadas no quarto experimento com a mesma metodologia do terceiro experimento. Os resultados encontrados, no entanto, foram mais interessantes que a previsão inicial. A discriminabilidade e a imitação corretas, esta última medida pelo VOT dos alofones em posição incorreta, não foram verificadas para os logatomas, apenas para as palavras existentes no léxico. A conclusão mostra-se clara: alofones podem ser percebidos pelos falantes, desde que inseridos em palavras da língua, o que mostra um forte efeito lexical na percepção de fala, já encontrado anteriormente para outros fenômenos de percepção de fala (e.g. Samuel (1987) para a restauração fonêmica e capítulo 3).

Os resultados encontrados pelos autores fornecem um indício inicial de que a hipótese da alofonia especificada lexicalmente possa ser verdadeira, talvez principalmente pelo fato de o fenômeno não se aplicar indistintamente, mas apenas onde a alofonia ocorre

²A e B são categorias diferentes e X pertence a uma das categorias, podendo ser, no entanto, um exemplar diferente de A e B. Em nossos experimentos, X será ou A ou B.

mais naturalmente. Este tipo de percepção alofônica sugere que é plausível os falantes fazerem uso de representações alofônicas no léxico sem que isso se torne contraproducente, pois, tanto para a produção quanto para a percepção, categorias diferentes estão sendo utilizadas.

Os experimentos que serão apresentados foram inspirados no experimento de WBI, adaptando-os ao PB e aperfeiçoando a metodologia para obter dados mais objetivos e adequados aos nossos objetivos.

1.4 Alofonia de [a, ə]

A alofonia de [a, ə] foi escolhida principalmente por 1) ser influenciada pelo acento lexical, como a alofonia de [p, p^h] no inglês americano e 2) ser uma alofonia já descrita em termos acústicos para o PB (AQUINO, 1997).

A influência do acento lexical poderia ser vista como determinante único da alofonia, i.e., o alofone [ə], seria uma consequência automática da atonicidade máxima. Lembremos porém, que no PB, diferentemente do português europeu, no qual há uma tendência mais forte à redução de vogais átonas a *schwa*, a redução vocálica é variável nos contextos átonos. Por exemplo, um /e/ pré-tônico sofre uma redução muito pequena numa palavra como “bebê” (AQUINO, 1997). Já em um contexto pós-tônico final, há acentuada diferença entre a redução de /e/ na palavra “mate”, onde há redução máxima para [ɪ] ([ˈmatʃɪ]), em relação a uma redução menor na palavra “ópera” ([ˈɔpərə]). Em alguns dialetos, e.g. no estado de Pernambuco (ARAGÃO, 1999), há a possibilidade de pronúncia da vogal [ɛ] em contextos átonos, em palavras como “menino” ([mɛˈninu]), diferentemente do que ocorre nos dialetos da região sudeste, onde há variação entre [meˈninu] e [miˈninu]. Exemplos como estes sugerem que a redução vocálica possa estar sobre o controle lingüístico no PB, o que justifica a escolha da alofonia de [a, ə] como uma alofonia candidata a especificação lexical.

Em termos acústicos, o alofone átono [ə] apresenta, considerando todos os contextos consonantais possíveis em PB, um F₁ mais baixo em relação a [a] e um F₂ mais elevado (AQUINO, 1997). Essas duas características acústicas sugerem um menor abaixamento de mandíbula e maior avanço do corpo da língua de [ə] em relação a [a]. O [ə] também apresenta duração reduzida em relação a [a] (AQUINO, 1997). Assim como no caso do

/e/ pós-tônico, há uma diferença entre o [ɐ] pós-tônico não-final e final. O primeiro é mais fechado e posteriorizado que o segundo (AQUINO, 1997). Albano e colegas (ALBANO et al., 1998) mostram evidência de que a redução do [ɐ] pós-tônico final é mais acentuada em contextos prosodicamente fracos do que em contextos prosodicamente fortes.

Teríamos então três categorias ou um contínuo? Albano (2001) sugere que a pergunta correta neste caso seria: há valor distintivo entre um alofone reduzido e um pleno? A resposta tende a ser afirmativa, já que a distinção pleno/reduzido para a autora está na base de sua proposta de acentuação para o PB. As vogais plenas³ mais à direita no item lexical seriam elementos atratores do acento lexical e as vogais reduzidas os repulsores.

Valendo-nos da evidência apresentada acima em favor de uma aparente especificação não-automática da alofonia de [a, ɐ], consideraremos os alofones como tendendo a categóricos (já que talvez não seja possível eliminar a gradiência completamente) e passíveis de estarem representados lexicalmente. Isto é, há evidência de que têm uma realização acústica e articulatória distintas e eles precisam estar indicados na entrada lexical de cada palavra em que ocorrem (ALBANO, 2001a; KOCHETOV, 2002). Essa condição visa uma compatibilidade com a hipótese dinâmica da FAR e seus princípios, os quais eliminam transformações simbólicas. Considerando a hipótese em termos de explicação de uma teoria lingüística, ela mantém simples a Fonologia, ao permanecer com os dois princípios mais simples da FAR, mas carrega de complexidade a especificação lexical ao propor um maior nível de detalhamento para a alofonia. Modelos gerativos representacionais como a Teoria da Otimalidade propõe especificar lexicalmente a alofonia discreta (KENSTOWICZ, 1998), mas ainda não são capazes de incorporar o detalhe fonético explicitado pelo gesto articulatório.

Nosso objetivo é, através de experimentos perceptuais com a alofonia de [a, ɐ] (capítulos 2 e 4), fornecer evidência para a plausibilidade empírica da hipótese da alofonia lexicalizada, investigando como os falantes da língua percebem o contraste alofônico de [a, ɐ]. Por fim, faremos uma breve discussão de como a FAR poderia incorporar variáveis ditas de desempenho como frequência de ocorrência.

³Toda a série de vogais do PB, não apenas os alofones [a, ɐ].

Capítulo 2

Experimento 1: Efeitos de variáveis lexicais na percepção da alofonia

Para testar a hipótese da influência lexical e percepção da alofonia de [a, ɐ] planejamos um experimento piloto utilizando pares mínimos de palavras e logatomas. Os pares mínimos foram escolhidos para que existisse a condição de comparação na qual duas formas contrastassem principalmente pela diferença alofônica. Com a inclusão das formas contrastantes também avaliar a influência do acento lexical e duração segmental na percepção de [a, ɐ].

A tarefa escolhida para verificar um possível efeito lexical foi a discriminação no paradigma *ABX* com medição do tempo de reação (doravante TR). Cada sujeito deveria responder no menor tempo possível se a terceira palavra ouvida na série era igual à segunda ou à primeira. A tarefa de discriminação permitiria investigar se há uma diferença significativa do número de erros entre palavras e logatomas, indicando uma possível sensibilidade dos falantes ao contexto lexical. Já o TR seria uma variável adicional com o mesmo objetivo da tarefa de discriminação, mas que poderia indicar situações em que houvesse facilitação e conseqüente diferenciação na percepção das formas.

A escolha pelo paradigma *ABX*, ao invés de *AXB* (mais utilizado em experimentos psicofísicos), justifica-se pela natureza auditiva do estímulo e pelo nível de controle que desejávamos obter. A apresentação *AXB* poderia estimular uma resposta antecipada em relação ao fim da série. O problema poderia ser contornado marcando como início da contagem do TR o fim da segunda forma, mas os sujeitos poderiam eventualmente optar por fazer a decisão apenas após ouvirem a terceira forma. Soma-se a isso o fato de que

as durações de cada forma são naturalmente variáveis (ver 2.1.3), o que introduziria uma variação adicional para aqueles sujeitos que respondessem após a terceira forma. Para forçar uma estratégia uniforme optamos, então, por utilizar o paradigma *ABX*, ainda que com uma maior sobrecarga de memória que o método acarreta. Essa sobrecarga, no entanto, não parece ter afetado o número de respostas válidas, sendo muito baixo o número de respostas efetuadas após o limite máximo permitido: apenas 16 em 2400 respostas ou 0,006%. Por fim, vale mencionar que paradigma *ABX* já foi utilizado com sucesso anteriormente em estudos de discriminação auditiva como o de Dupoux e colegas (1999).

O uso das tarefas de discriminação e da medição de TR ao invés da tarefa de classificação é justificado pela necessidade de medidas mais automáticas em que efeitos pós-lexicais fossem diminuídos ao mínimo possível. Também não era possível prever em que medida as classificações sobre o grau de acento estrangeiro para o inglês americano e PB poderiam ser comparadas entre os falantes. Também por ser uma medida menos objetiva, possivelmente estaria sujeita a uma maior variação individual, o que não seria desejável para o *design* do experimento.

2.1 Metodologia

2.1.1 *Corpus* utilizado

O *corpus* utilizado para construir os estímulos foi obtido em gravações em uma cabine com tratamento acústico e produzido por um locutor experiente. A taxa de elocução foi monitorada com o auxílio de um metrônomo, sendo repetidas as locuções não-ótimas do ponto de vista da pronúncia e da manutenção da taxa de elocução estatisticamente constante. Foram utilizadas palavras e logatomas que propiciassem as ocorrências de todas as vogais desejadas num contexto consonantal e acentual controlados (apêndice 5.3). Foram gravados dois contextos de ocorrência: um no qual a vogal estava em final absoluto de oração: “Digo *palavra/logatoma*” e outro em que estava no meio de uma frase veículo: “Digo *palavra/logatoma* rápido.” Apenas a primeira condição foi utilizada na geração dos estímulos, já que esta seria a condição de ocorrência das palavras no experimento.

2.1.2 Sujeitos

Neste primeiro momento, o experimento foi realizado com dez sujeitos, todos universitários, com idade em torno de 20 anos e sem problemas de audição declarados. Nenhum sujeito executou a tarefa com dificuldade maior que a esperada ou mencionou qualquer desconforto durante a execução do teste. Os dados de todos os sujeitos foram incluídos na análise estatística.

2.1.3 Estímulos

Neste primeiro experimento escolhemos o verbo *pagar* no infinitivo e o logatoma *dába* com as respectivas formas contrastantes *paga* e *dabá*. Sendo o verbo *pagar* um verbo de alta frequência em PB, adotamos pronúncia [pa'gɐ], que tende a ser mais realizada na fala cotidiana no lugar de [pa'gaɫ], que tende a ser realizada apenas em fala hiper-articulada ou formal.

Os estímulos foram obtidos concatenando-se os alofones produzidos em posição natural de ocorrência em um ambiente incorreto, i.e., numa posição em que não ocorrem naturalmente. Por exemplo, para obtermos a forma [pɐgɐ], retiramos o segundo /a/ do contexto “Digo [papegɐ]”¹ e o substituímos na primeira posição da forma [pɐgɐ].

Para a concatenação, utilizamos uma técnica de síntese de fala híbrida conhecida como H+N (*Harmonic plus Noise*, desenvolvida por Violaro e Böeffard (1998)). A síntese H+N, além de permitir a concatenação suave dos trechos, nos permite manipular F_0 e duração independentemente, possibilitando um maior controle e manipulação destas variáveis. As figuras 2.1 e 2.2 mostram, respectivamente, espectrogramas da forma concatenada [pɐ'ga] e da forma [pa'ga] sem concatenação.

A partir do conjunto de palavras obtivemos as seguintes formas:

1. A partir da forma-base oxítona: [pa'ga], [pɐ'ga] e [pa'gɐ];
2. A partir da forma-base paroxítona: [pɐgɐ], [pɐgɐɐ] e [paga].

E a partir dos logatomas:

1. A partir da forma-base oxítona: [da'ba], [dɐ'ba] e [da'bɐ];

¹Neste primeiro experimento utilizamos vogais pós-tônicas em contexto final e não-final (e.g. [paga]). No segundo experimento, uniformizamos o contexto, utilizando apenas formas finais (e.g. [pape 'gatɐ] ao invés de [papegɐ], seção 4.1.3).

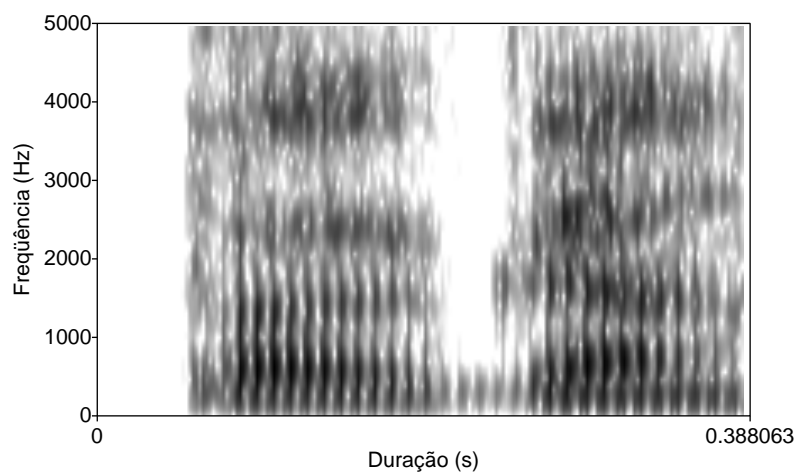


Figura 2.1: Espectrograma de [pɐ'ga] concatenado, experimento 1.

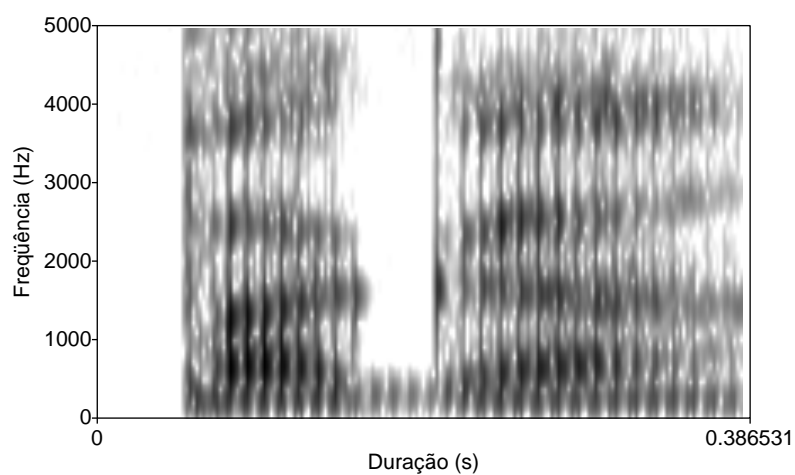


Figura 2.2: Espectrograma de [pa'ga] sem concatenação, experimento 1.

2. A partir da forma-base paroxítona: ['dabə], ['dəbə] e ['daba].

Tanto para as palavras quanto para os logatomas do *corpus* temos, então, três variantes por posição acentual (tônica e pós-tônica), totalizando doze estímulos diferentes.

2.1.4 Variáveis

As variáveis independentes investigadas no experimento 1 são:

1. Combinação das formas de palavras e logatomas nas tríades (seção 2.1.3):

ABA e ABB;

ACA e ACC;

BCB e BCC. Em que, por exemplo, A = ['pagə], B = ['pəgə] e C = ['paga].

2. Lexicalidade, i.e., palavras \times logatomas;

3. Duas condições acentuais distintas:

acento paroxítono;

acento oxítono.

4. Dois padrões de duração para os segmentos:

Vogais e consoantes com duração fixa ajustada pela média global da palavra/logatoma correspondente em ambas posições acentuais;

Vogais e consoantes com a duração variável, determinada pela média aritmética de cada posição acentual obtida no *corpus* natural gravado (seção 2.1.1).

E como variáveis dependentes:

1. Tempo de reação;

2. Acertos/erros.

O F_0 e a amplitude² do sinal foram apenas controlados, sendo mantidos constantes entre os estímulos, para que tivéssemos um número de variáveis manipuladas relativamente pequeno e conseqüentemente um *design* experimental gerenciável para o primeiro

²A amplitude do sinal foi manipulada no segundo experimento (capítulo 4.)

experimento. No caso da amplitude, isso gerou estímulos mais artificiais, pois ambas as vogais sempre apresentavam nível de intensidade similar. Essa artificialidade poderia se confundir com a influência dos fatores manipulados. No entanto, os resultados do segundo experimento indicam que a intensidade parece não influenciar significativamente a percepção da alofonia de [a, ɐ] (seção 4.4.1).

2.1.5 Procedimentos

O experimento foi realizado em uma sala silenciosa, com os sujeitos sentados em frente a um monitor e ouvindo os estímulos através de um fone de ouvido. A amplitude do sinal foi ajustada para um nível confortável e mantida constante para todos os sujeitos. Antes do início do experimento, as instruções (ver apêndice 5.3) foram apresentadas no monitor e eventualmente esclarecidas pelo experimentador, que permanecia na sala durante toda a duração do teste, aproximadamente vinte e cinco minutos. Para obter as respostas utilizamos um *joystick* de microcomputador do fabricante *Genius*, modelo *Max-Fire Digital G-09D*. O *software* utilizado na apresentação dos estímulos foi o DMDX versão 3 (FORSTER, K. I.; FORSTER, J. C., 2003). Tanto o *joystick* como mecanismo de entrada quanto o *software* propiciaram tempos de reação com precisão de milissegundo.

O experimento foi dividido em oito blocos de apresentação, com pequenas pausas opcionais após cada bloco e uma pausa maior após o quarto bloco. Em cada bloco foram apresentadas 30 tríades: 3 estímulos distintos \times 10 repetições. As condições acentuais foram apresentadas em blocos distintos. No total cada sujeito respondeu 240 tríades. A apresentações dos blocos e das tríades no interior de cada bloco foi aleatória para cada sujeito. Antes do início do experimento, foi apresentado um conjunto de treino contendo doze tríades a serem discriminadas, não sendo computados estes resultados. Os intervalos entre cada palavra/logatoma nas tríades era de 200 ms e entre tríades de 2 s. O intervalo máximo para cada resposta foi de 2 s.

2.2 Análise Estatística

A análise estatística do experimento consistiu em dois tipos de análise separadas:

- χ^2 (Chi-quadrado) de *Pearson* para os dados de acertos/erros;
- Análises de variância (ANOVA) com medidas repetidas para os dados de TR.

2.2.1 Análise Estatística do tempo de reação

Preliminares

A análise do TR incluiu apenas os dados das respostas corretas, i.e., aquelas em que o sujeito associou corretamente a terceira forma com a primeira ou com a segunda forma. A inclusão do TR das respostas incorretas introduziria uma variação adicional nos dados, já que não é possível distinguir as situações em que o erro é resultado da dificuldade da tarefa daquelas em que o erro é um engano cometido pelo sujeito no momento da resposta. Os erros no total, tiveram sua frequência distribuída entre os blocos de resposta e entre os sujeitos. A tabela 2.1 mostra o número médio de erros por bloco e sujeito. O erro médio para os sujeitos é o total de erros de todos os sujeitos dividido pelo número de sujeitos. Já o erro médio para os blocos foi calculado dividindo-se o total de erros em cada bloco pelo número de sujeitos. O mínimo e máximo para os erros dos blocos são os extremos de erros cometidos por um sujeito dentro de um bloco experimental.

	erro médio	σ	mínimo	máximo	erro geral %
Blocos (8)	3,9	0,8	0	11	13%
Sujeitos (10)	31,2	16,7	6	57	13%

Tabela 2.1: Padrão de erro médio por bloco para os blocos e sujeitos, experimento 1.

O objetivo da análise estatística de TR foi testar se há diferentes padrões de resposta quando há variação nos fatores independentes (seção 2.1.4).

A primeira dificuldade na análise do TR foi decidir em que medida métodos estatísticos paramétricos eram adequados aos dados de TR. Tais dados têm uma tendência natural de apresentar uma distribuição com assimetria (*skewness*) com cauda à direita. A figura 2.3 mostra o histograma e a figura 2.4 o gráfico de probabilidade normal para os dados de todos os sujeitos. A não-normalidade dos dados também é reportada em estudos de TR de movimentos de sacada ocular (CARPENTER; WILLIAMS, 1995). Desvios de normalidade como os apresentados podem comprometer os resultados das ANOVAs, pois os testes assumem a normalidade dos dados. Por outro lado, utilizar testes não-paramétricos tornaria a análise extremamente complexa, especialmente pela necessidade de utilizar análises de medidas repetidas, e com menor capacidade de detectar diferenças estatisticamente significativas.

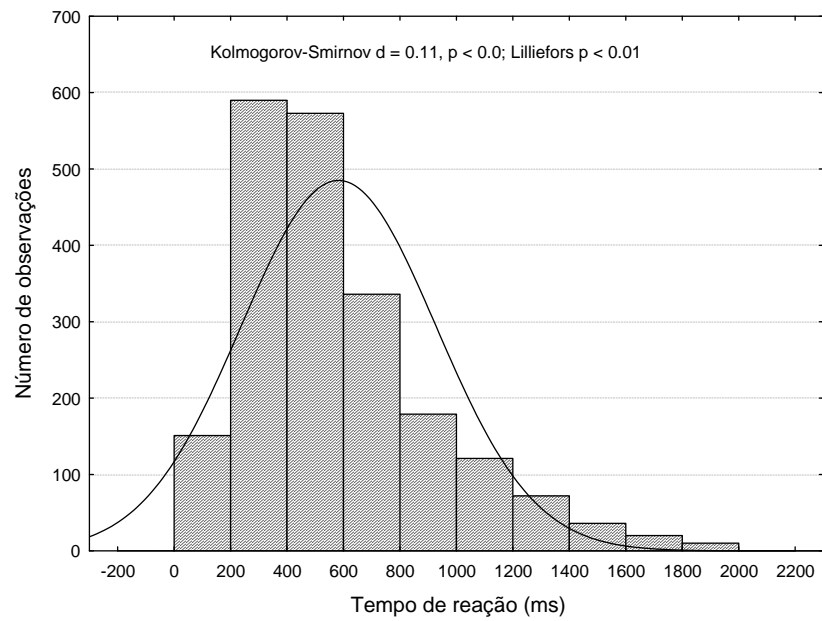


Figura 2.3: Histograma do tempo de reação do experimento 1 com dados para todos os sujeitos e resultados dos teste de normalidade *Kolmogorov-Smirnov* e *Lilliefors*.

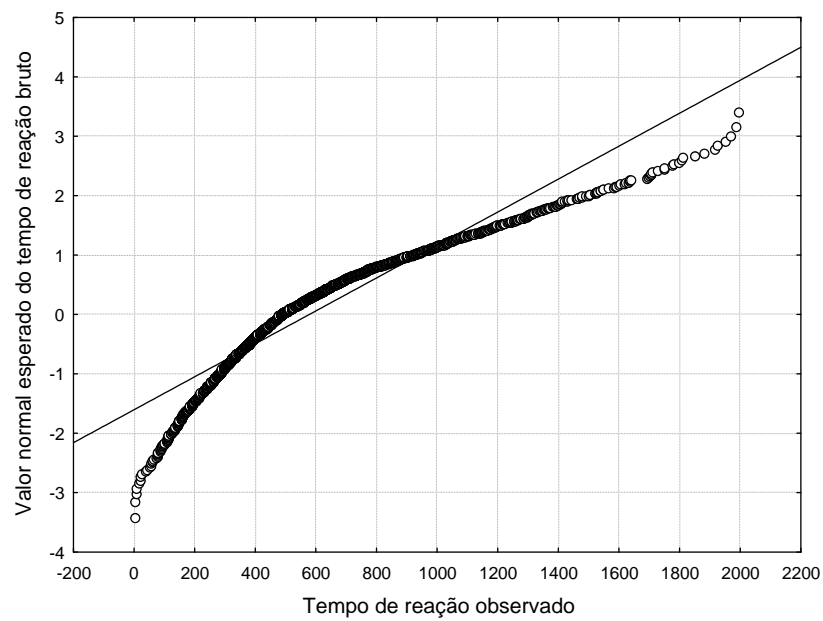


Figura 2.4: Gráfico de probabilidade normal do experimento 1 com dados para todos os sujeitos.

Uma segunda dificuldade era sobre como tratar os possíveis *outliers* dos dados de TR. Ainda que apenas para as respostas corretas, os dados apresentavam ocasionalmente valores de TR abaixo de 50 ms e acima de 1800 ms (com média de 578 ms e mediana de 486 ms) e que poderiam ser *outliers*. A solução mais usual de eliminar da análise os possíveis *outliers* incorreria em dois problemas: 1) a não-normalidade dos dados e conseqüente impossibilidade de aplicar a maior parte das técnicas de eliminação de outliers; e 2) a impossibilidade de determinar o que seriam os *outliers*, já que eles poderiam ter várias causas³ como falha motora, desatenção, dificuldade na decisão, antecipação na resposta, etc.

Normalização do tempo de reação

A solução adotada para contornar tanto a não-normalidade dos dados quanto a presença de possíveis *outliers* foi a normalização, i.e., tornar a distribuição dos dados mais próxima de uma distribuição normal. Para tal, utilizamos a transformação *Box Cox* (BOX; COX, 1964), definida pela equação 2.1:

$$y = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \ln x & \text{se } \lambda = 0 \end{cases} \quad (x > 0) \quad (2.1)$$

em que x é o valor a ser transformado, λ é um parâmetro a ser determinado empiricamente e $\ln x$ é o logaritmo natural do valor a ser transformado.

A transformação *Box Cox* é apropriada exatamente nos casos em que há alguma assimetria na distribuição. Para os nossos dados, temos $\lambda = 0,295$. Calculamos λ por tentativa e erro, de modo que os valores da assimetria e curtose do dados normalizados se aproximassem de zero. Na tabela 2.2 temos um sumário estatístico dos dados de TR com e sem normalização⁴. Fica evidente que a transformação aplicada reduz a não-normalidade dos dados⁵, ainda que, com ou sem normalização, os dados não sejam estritamente normais em termos estatísticos, pois os testes de normalidade mais comuns (*Kolmogorov-Smirnov*, *Lilliefors* e *Shapiro-Wilk*) costumam ser sensíveis a *outliers* em grandes amostras (maiores

³A possibilidade de outliers decorrentes de erro na entrada de dados nas planilhas para a análise estatística está eliminada já que todos os passos foram automatizados com o auxílio de programas de manipulação de dados desenvolvidos para esta finalidade.

⁴Valores calculados apenas com as respostas corretas, ainda sem incluir as médias que preencheram os dados faltantes para que fosse possível realizar a análise estatística.

⁵Lembrando que a assimetria e a curtose da distribuição normal teórica é zero e média e mediana têm o mesmo valor.

	média	mediana	σ	faixa	assimetria	curtose
TR	578	486	343	1993	1,22	1,47
TR $\lambda = 0,295$	17,9	17,6	3,8	30,1	-0,04	0,7

Tabela 2.2: Valores de média, mediana, desvio padrão, faixa de variação (em ms), assimetria e curtose para os tempos de reação brutos e normalizados, experimento 1.

que 1000 observações). As figuras 2.5 e 2.6 mostram, respectivamente, o histograma e o gráfico de probabilidade normal dos dados de TR normalizados.

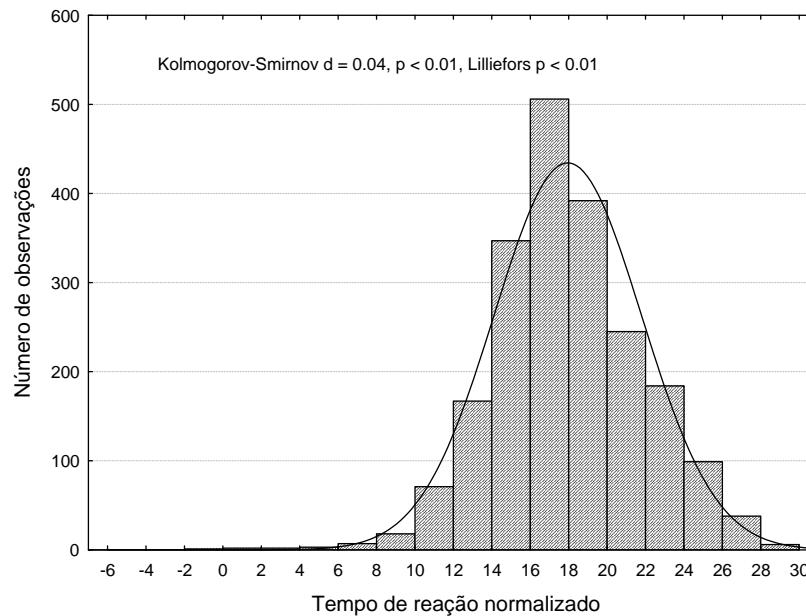


Figura 2.5: Histograma do tempo de reação do experimento 1 com dados normalizados para todos os sujeitos.

Uma possível objeção ao nosso procedimento de normalização é a falta de motivação empírica para aplicar uma transformação não-linear nos dados. A nossa motivação é metodológica, mas para nos certificar de que a transformação não introduziria transformações que pudessem alterar os resultados dos testes estatísticos, realizamos os testes estatísticos também com os dados não transformados e em nenhum caso houve alteração na rejeição ou confirmação das hipóteses nulas. Em nosso caso, a ANOVA mostrou-se robusta ao desvio de normalidade apresentado em nossos dados, em consonância com as simulações de Monte Carlo que mostram a robustez das ANOVAs a alguns desvios de normalidade (GLASS; PECKHAM; SANDERS, 1972). A figura 2.7 mostra que a a faixa que sofre transformação mais acentuada é justamente a das caudas da distribuição, onde se encontram, possivelmente, a maioria dos *outliers*.

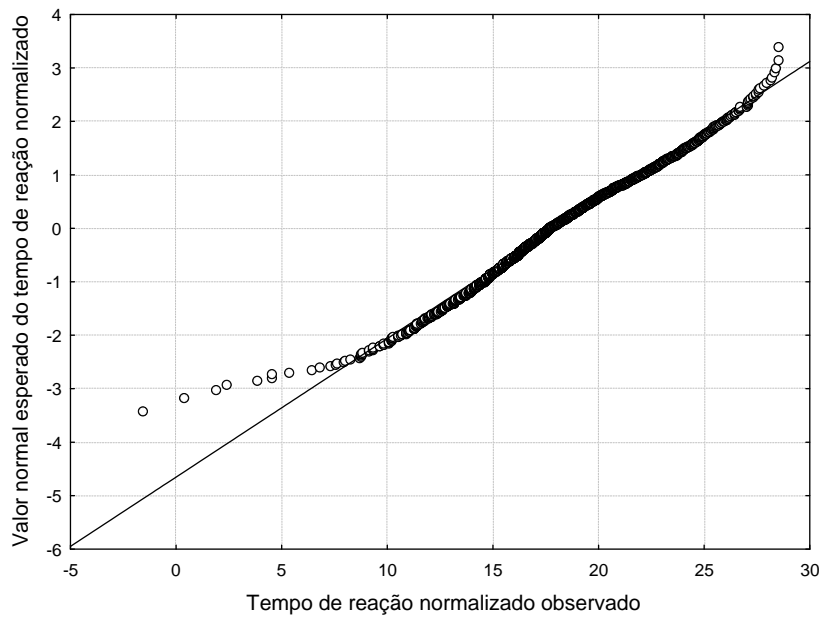


Figura 2.6: Gráfico de probabilidade normal do experimento 1 com dados normalizados para todos os sujeitos.

A normalização dos dados permitiu um maior rigor na aplicação dos testes de hipótese paramétricos sem eliminar dados arbitrariamente. Reportaremos apenas os dados das análises estatísticas com os dados normalizados.

Análise de variância com medidas repetidas

O teste estatístico paramétrico mais apropriado aos nossos dados é a Análise de Variância com medidas repetidas. O teste com medidas repetidas é utilizado quando mais de uma medição é realizada na mesma unidade experimental ou em medidas em sucessão temporal ou em ambos os casos. No nosso experimento, ambas as situações se aplicam: temos mais de uma medida por sujeito (potencialmente 240 válidas), que é a nossa unidade experimental, pois há variação de condições experimentais (fatores independentes), e dentro da mesma condição, temos a sucessão temporal, no caso das dez repetições da mesma tríade dentro de um bloco de apresentação (seção 2.1.5). Em nossos experimentos é irrelevante a regularidade da sucessão temporal, já que as repetições de uma mesma condição eram apresentadas em ordem aleatória e não detectamos nenhuma diferença significativa de tempo de reação entre as repetições ($F(4, 6) = 2,72$, $p = 0,13$), i.e., não parece haver um efeito de aprendizado na tarefa com a apresentação aleatória.

Um dos requerimentos da ANOVA com medidas repetidas é que não haja dados faltantes (*missing*). Como descartamos as respostas incorretas e os *timeouts* das análises,

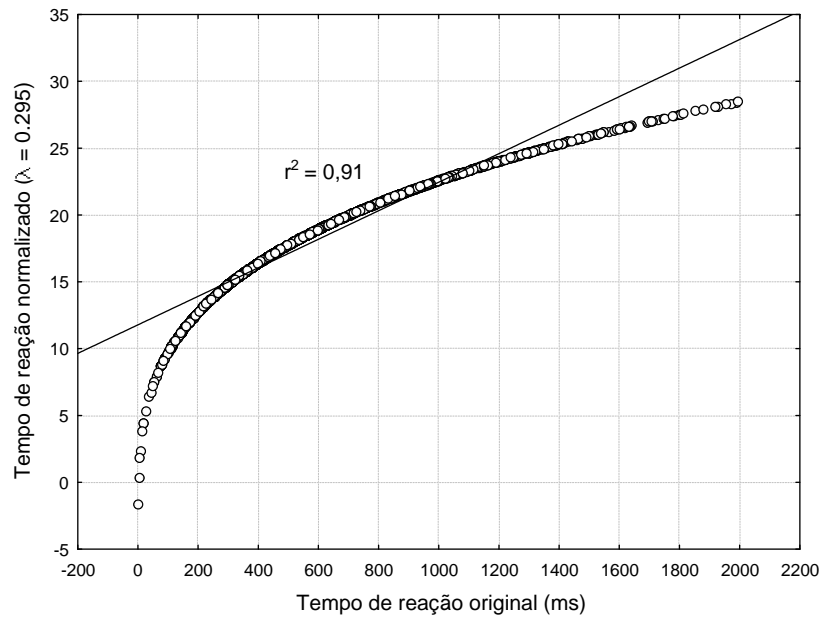


Figura 2.7: Correlação entre o tempo de reação normalizado e o tempo de reação bruto para todos os sujeitos e resultados dos teste de normalidade *Kolmogorov-Smirnov* e *Lilliefors*.

não conseguiríamos os dados completos para nenhum sujeito se desejássemos analisar os dados com as dez repetições. A solução encontrada foi utilizar apenas cinco repetições aleatórias, completando com as médias dos sujeitos os casos em que não houvesse nem cinco respostas válidas por condição, num total de 13% do total. Com esse procedimento, tentamos tornar a análise possível, introduzindo um mínimo necessário de tendência central nos dados.

Ordenamos os dados de acordo com as recomendações de Max e Onghena (1999) para a análise de medidas repetidas em experimentos lingüísticos, considerando cada sujeito como uma única unidade experimental. Assim, obtivemos 120 condições consideradas como variáveis dependentes na análise, a saber: 5 repetições \times 3 combinações entre as tríades ABC \times 2 condições de lexicalidade \times 2 condições acentuais \times 2 condições de duração.

2.2.2 Análise Estatística dos acertos/erros

Para detectar uma possível diferença no padrão de acertos/erros na discriminação das tríades a depender do fator independente, realizamos o teste de χ^2 (chi-quadrado) de *Pearson* para comparar todos os fatores independentes. No caso do fator que representa as combinações das tríades, separamos as tríades por número de formas incorretas (em

cada tríade há de uma a três formas incorretas) e realizamos as comparações agrupando duas condições por análise, num total de três análises distintas.

2.3 Resultados

2.3.1 Resultados de tempo de reação

A análise de variância com medidas repetidas apontou um resultado estatisticamente significativo apenas para o fator acento lexical ($F(1, 9) = 49,314, p < 0.001$), não envolvendo problemas de esfericidade por se tratar de fator com apenas dois níveis (MAX; ONGHENA, 1999). Não houve diferença significativa para os outros fatores ou para qualquer interação entre os fatores.

Os sujeitos testados detectaram mais rapidamente a diferença entre os alofones quando estes se encontravam em sílabas paroxítonas, o padrão acentual canônico do português.

2.3.2 Resultados de padrão de acertos/erros

Para o padrão de acertos/erros, apenas o fator lexicalidade apresentou uma distinção estatisticamente significativa ($\chi^2(1) = 9,21, p = 0,0024$). Para o fator acento lexical, houve uma diferença estatística marginal ($\chi^2(1) = 3,77, p = 0,052$), na direção de um maior número de erros para as formas oxítonas. Os resultados sugerem que os sujeitos parecem ter uma dificuldade maior em distinguir entre os alofones [a, ɐ] quando estes se encontram em palavras, mas esta dificuldade não é clara quando a condição de padrão acentual é variada. O fator padrão de duração não apresentou diferença estatisticamente significativa. A taxa de erro médio ficou em 13% do total de discriminações. A tabela 2.3 resume o padrão de respostas corretas e incorretas para os fatores independentes binários (lexicalidade, padrão acentual e padrão de duração).

Para a análise das combinações de tríades, realizamos três análises separadas:

1. Tríades com uma forma incorreta \times tríades com duas formas incorretas: $\chi^2(1) = 21,93, p < 0,001$;
2. Tríades com uma forma incorreta \times tríades com três formas incorretas: $\chi^2(1) = 65,69, p < 0,001$;

	acertos	erros
palavras	1019	181
logatomas	1069	131
paroxítonos	1060	140
oxítonos	1028	172
duração constante	1045	155
duração variável	1043	157

Tabela 2.3: Padrão de acertos/erros para os fatores independentes binários, experimento piloto.

3. Tríades com duas formas incorretas \times tríades com três formas incorretas: $\chi^2(1) = 13,13$, $p < 0,001$;

Em todas as comparações, há diferenças significativas no número de acertos e erros a depender da quantidade de estímulos incorretos presentes nas tríades. Há maior número de erros quanto menor o número de formas incorretas, ou seja, há uma maior exatidão na discriminação dos alofones quando as formas incorretas predominam. A figura 2.8 exibe graficamente os resultados destes testes estatísticos.

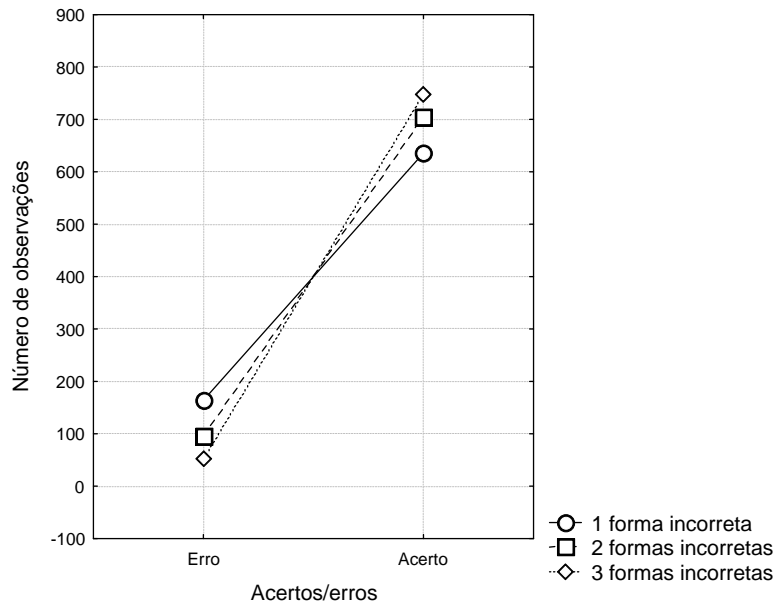


Figura 2.8: Quantidade de acertos/erros para as tríades classificadas por quantidade de formas incorretas, experimento 1.

2.4 Discussão

Os resultados da análise do tempo de reação (seção 2.3.1) indicam que há uma diferença temporal na decisão apenas quando há variação da posição acentual.

Já os resultados para o padrão de acertos/erros sugere que o fator lexicalidade é que diferencia a acuidade da discriminação, havendo uma diferenciação possível (resultado marginal) para o fator padrão acentual. Os dois tipos de resultados indicaram neste primeiro momento que é plausível a hipótese de uma diferenciação na percepção da alofonia de [a, ɐ] a depender do fator lexicalidade (e conseqüentemente distinção por frequência de ocorrência, já que logatomas bem formados podem ser vistos como palavras de baixíssima frequência). Há uma diferença, num nível acima do aleatório, de como os sujeitos discriminam alofones em posição incorreta em palavras e logatomas. Já a ausência de quaisquer efeitos significativos para os diferentes padrões de duração em ambos os tipos de resposta sugere que a duração talvez seja um fator pouco relevante para a percepção da alofonia de [a, ɐ] em posição de final de palavra, contexto no qual há um alongamento natural da vogal final.

No entanto, a falta de concordância entre os resultados de acertos/erros e de TR nos levou a considerar que um maior nível de controle das variáveis que poderiam influenciar na tarefa de discriminação era desejável. Investigamos a densidade de vizinhança fonológica (seção 3.2) das palavras e logatomas utilizados neste primeiro experimento e verificamos que havia uma diferença não controlada e que poderia influenciar em nossos resultados, um efeito já detectado para o inglês (LUCE; PISONI, 1998). A figura 2.9 mostra os pares de palavras e logatomas e suas respectivas vizinhanças fonológicas.

Desenvolvemos então um segundo experimento (capítulo 4), no qual controlamos minimamente vizinhança fonológica e fonotaxe, que discutiremos no próximo capítulo, e aperfeiçoamos alguns detalhes experimentais.

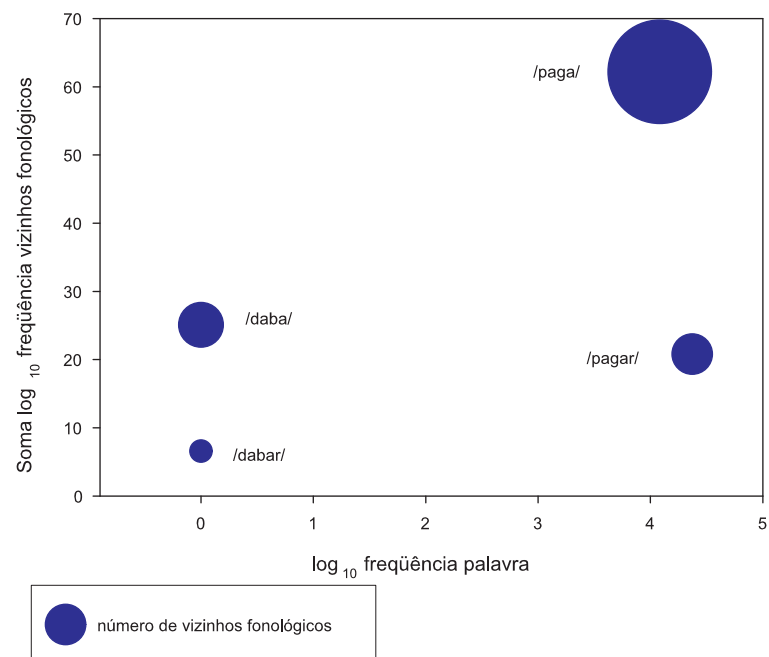


Figura 2.9: Frequência de ocorrência dos pares /paga paga(r)/ e /daba daba(r)/, medida de vizinhança fonológica e número absoluto de vizinhos fonológicos (ver capítulo 3).

Capítulo 3

Frequência de ocorrência, vizinhança lexical e fonotaxe

3.1 Frequência de ocorrência das palavras

Para a tradição lingüística gerativista, a frequência do uso das palavras na língua sempre foi vista como um problema de performance, não sendo parte da gramática de uma língua. Dentro da teoria lingüística, a área de maior destaque em que a frequência de ocorrência das palavras na língua apresenta-se como relevante é a Sociolingüística. Esta, entretanto, usualmente não estuda a relação entre frequência de uso e conhecimento gramatical, mas a relação da frequência de uso associada a fatores extra-lingüísticos.

Já no campo da Psicologia Experimental, o efeito da frequência de uso das palavras (ainda que rotineiramente estimado por um *corpus* de texto escrito) possui uma longa tradição de estudos empíricos (WHALEY, 1978; HOWES; SOLOMAN, 1951). Resumidamente, palavras mais frequentes na língua são detectadas, discriminadas e repetidas mais rapidamente do que palavras raras em testes como decisão lexical e nomeação (*naming*), tanto auditivos quanto de leitura¹. A figura 3.1 mostra a correlação entre o logaritmo de base 10 da frequência da palavra (doravante $\log_{10} f$) e a média dos tempos de reação para todos os sujeitos num teste de decisão lexical a partir de leitura para 180 palavras do PB (FRANÇOZO et al., 2002).

¹A decisão lexical consiste em responder se uma palavra apresentada pertence ou não à língua. Já o que denominamos nomeação envolve pronunciar uma forma ouvida ou lida.

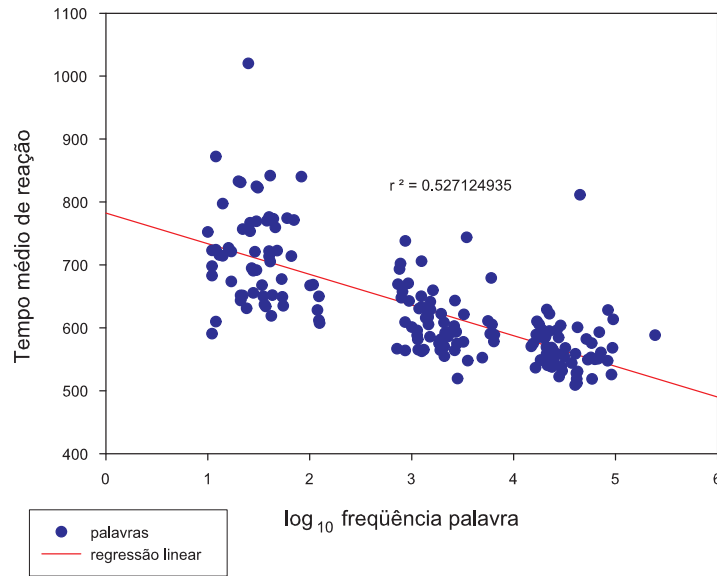


Figura 3.1: Correlação entre $\log_{10} f$ e média dos tempos de reação para 180 palavras do PB numa tarefa de decisão lexical a partir de leitura.

É importante mencionar que a frequência de ocorrência é geralmente representada em logaritmo devido à natureza da distribuição da ocorrência de palavras numa língua. Se ordenarmos graficamente a ordem de frequência (*ranking*) das palavras pela sua frequência absoluta em eixos logarítmicos, teremos pontos que se aproximarão de uma reta com inclinação negativa e aproximados pela relação expressa na equação 3.1.

$$\mathbf{F}_n \sim \frac{1}{n^\alpha} \quad (3.1)$$

em que \mathbf{F}_n é a frequência de ocorrência da n ésima palavra e α é uma variável determinada empiricamente e próxima a 1.

Esta distribuição, primeiramente apresentada por George Zipf (1949) e conhecida por Lei de Zipf, parece ser válida para diferentes tipos de *corpora*: escrita, fala ou a obra de um único autor. A figura 3.2 mostra a lei de Zipf para as 20 mil palavras mais frequentes do CETENFolha (seção 4.1.1).

Esse efeito da frequência, apesar de ser um dos efeitos mais robustos em experimentos psicofísicos, não foi incorporado a modelos lingüísticos de percepção de fala. Modelos como a Teoria Motora da Percepção (LIBERMAN; MATTINGLY, 1985), a Teoria Quântica da fala (STEVENS, 1989) e o Realismo Direto (FOWLER, 1986), para citar alguns dos mais relevantes, não incorporam a informação de frequência de ocorrência. É claro que, com

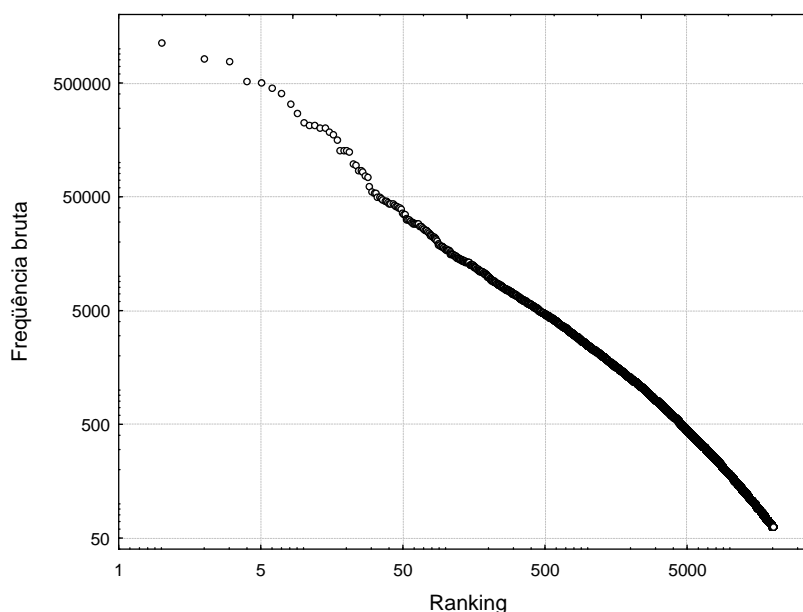


Figura 3.2: Lei de *Zipf* para as 20 mil palavras mais freqüentes do CETENFolha.

exceção do Realismo Direto, todos estes modelos aderem à teoria gerativa da gramática, para a qual a distinção entre competência e performance justifica a exclusão de fatores ditos de desempenho na construção do modelo. No entanto, as evidências da influência do efeito de freqüência na percepção de fala, a qual é usualmente considerada parte da gramática, levantam dúvidas se essa separação rígida é uma simplificação aceitável.

Pelo lado dos estudos em que há predomínio da tradição da Psicologia, há modelos de percepção de fala que incorporam informação de freqüência e até de vizinhança fonológica (ver 3.2), como o *Neighborhood Activation Model* (LUCE; PISONI, 1998) e o *Merge* (NORRIS; MCQUEEN; CUTLER, 2000). Ambos, entretanto, são modelos de reconhecimento de palavras isoladas, articulando minimamente a percepção de fala com outros componentes da gramática. Ambos utilizam o segmento como unidade fonológica mínima, deixando de lado os ganhos da representação do segmento em unidades menores como traços articulatórios e também deixando de incorporar a dimensão temporal do gesto articulatório.

E como conciliar os efeitos de freqüência com teorias lingüísticas de percepção (e possivelmente produção) de fala? O primeiro passo é esvaziar a distinção entre competência e desempenho da Gramática Gerativa². Esta divisão ainda é bastante tradicional e assu-

²É claro que isso não implica necessariamente em rejeitar completamente a possibilidade de determinados requisitos para a linguagem serem inatos, apenas significa uma menor ênfase no estudo de aspectos inatos.

mida dentro da Lingüística, dado o sucesso da Gramática Gerativa em explicar e prever os padrões sintáticos das línguas. Não é nosso objetivo aqui discutir o mérito da proposta inatista, mas observar que uma teoria de percepção e produção de fala que incorpore fatores que não são puramente gramaticais terá necessariamente de reconsiderar os limites da divisão entre competência e desempenho tais como usualmente apresentados.

O segundo passo seria incorporar informação estatística e probabilística em modelos de produção e percepção de fala. Nesta linha de investigação, tanto para a percepção quanto para a produção de fala, destacamos os trabalhos de Pierrehumbert (2001; 2003), Frisch (1996) e Albano (1999; 2001). Todos os autores propõem uma diferenciação em relação ao modelo gerativista ao incorporarem distribuições estatísticas e probabilísticas em busca de modelos com maior predição de aspectos não-categóricos das línguas. A diferença entre eles é, principalmente, a unidade fonológica mínima adotada. Frisch e Pierrehumbert utilizam traços articulatórios e Albano o gesto articulatório. Os trabalhos de Frisch e Pierrehumbert enfatizam que, quando métricas quantitativas são consideradas como um modelo de explicação gramatical, elas conseguem explicações mais completas que abordagens puramente categóricas. O custo é, obviamente, uma maior complexidade da descrição e do modelo gramatical. Como exemplo desse tipo de evidência, o trabalho de Frisch (1996) mostra que a frequência de ocorrência de segmentos, devidamente decompostos em traços articulatórios, tem um papel importante na explicação da fonotaxe do inglês e do árabe.

3.2 Vizinhança fonológica

Antes de abordar a importância da vizinhança lexical, cabe definir o conceito de vizinhança fonológica segmental que adotamos, que é o conceito clássico de Coltheart (1977):

Definição 1 *Uma palavra x é vizinho fonológico de uma palavra y quando x difere em apenas um fone de y por substituição, adição ou apagamento.*

Exemplificando: *fraca*, *facas* e *maca* são vizinhos fonológicos da palavra *faca*. O conceito de vizinhança fonológica de uma palavra é uma possível medida de grau de similaridade fonológica segmental entre formas.

É claro que como medida de similaridade fonológica, a vizinhança fonológica puramente segmental é uma aproximação bastante simplificada. Assume-se que a diferença

entre dois segmentos quaisquer é sempre a mesma, independente dos segmentos comparados. Por exemplo, a distância entre [p] e [k] é a mesma de [p] e [b]. A métrica desejável para nosso estudo seria uma métrica baseada em gestos articulatórios, pois poderia capturar os contrastes lexicais na mesma unidade básica do nosso ponto de partida. Entretanto, essa métrica ainda está em fase inicial de desenvolvimento no LAFAPE por Plínio A. Barbosa e depende ainda de simulações e medições articulatórias acuradas. Uma métrica alternativa para vizinhança fonológica é a desenvolvida por Frisch (1996), que se baseia em traços distintivos articulatórios. Apesar da vantagem de diferenciar a influência dos diferentes segmentos, optamos por manter a definição tradicional, dados os problemas inerentes de representação dos traços articulatórios em suas diferentes versões (D'ANGELIS, 1998; SILVA, 2002) e da acentuada incompatibilidade dos traços articulatórios com uma representação gestual. A **Definição 1**, ao expressar a vizinhança em termos segmentais, tem a vantagem sobre os traços de capturar as relações mais globais entre os gestos, se considerarmos o segmento como uma cristalização no léxico de padrões gestuais estáveis (LÖFQVIST, 1997), incluindo também a especificação laríngea destes gestos.

Os dados da figura 3.1 mostram que a relação entre $\log_{10} f$ e tempo de reação está longe de ser perfeita. Ainda que análises de variância com a frequência de ocorrência das palavras categorizada em três níveis nos mostre uma distinção, uma parte da variância dos dados permanece inexplicada (cerca de 48%), sendo que parte dessa variância inclui diferenças individuais entre os sujeitos.

Um modelo de reconhecimento de palavras isoladas bastante completo, fundamentado em evidências empíricas e que leva em conta a vizinhança fonológica é o já mencionado *Neighborhood Activation Model* ((LUCE; PISONI, 1998), doravante apenas NAM). Além da frequência dos itens lexicais, o NAM inclui o número de vizinhos fonológicos (densidade da vizinhança), a frequência destes vizinhos (frequência da vizinhança) e um índice de inteligibilidade entre palavras monossilábicas do inglês americano.

As frequências das palavras no NAM foram definidas a partir de um *corpus* de texto escrito (KUCERA; FRANCIS, 1967) e a vizinhança foi estabelecida de acordo com a **Definição 1** com o auxílio de um *corpus* computadorizado. Já o índice de inteligibilidade foi obtido a partir de um experimento prévio em que falantes do inglês identificavam por escrito monossílabos em diferentes níveis de interferência de ruído branco. Com

os resultados das identificações estabeleceu-se um índice de acerto para cada palavra, posteriormente usado para determinar o quanto uma determinada palavra poderia ser confundida com seus vizinhos fonológicos.

O resultado é uma métrica que tenta sumarizar as influências conjuntas da frequência e da vizinhança fonológica. A equação que representa esta métrica, denominada **Frequency-Weighted Neighborhood Probability Rule (FWNPR)** é:

$$\frac{\prod_{i=1}^n p(\mathbf{PS}_i|\mathbf{PS}_i) * \mathbf{Freq}_s}{\{[\prod_{i=1}^n p(\mathbf{PS}_i|\mathbf{PS}_i)] * \mathbf{Freq}_s\} + \sum_{j=1}^{nn} \{[\prod_{i=1}^n p(\mathbf{PN}_{ij}|\mathbf{PS}_i)] * \mathbf{Freq}_{Nj}\}} \quad (3.2)$$

Em que \mathbf{PS}_i é a probabilidade do i ésimo fonema da palavra, $p(\mathbf{PS}_i|\mathbf{PS}_i)$ é a probabilidade condicional de identificar o i ésimo fonema da palavra dado o mesmo fonema, \mathbf{PN}_{ij} é a probabilidade do i ésimo fonema do j ésimo vizinho, n é o número de fonemas na palavra e no seu vizinho, \mathbf{Freq}_s é a frequência da palavra, \mathbf{Freq}_{Nj} é a frequência do j ésimo vizinho e nn é o número de vizinhos da palavra. Os termos de probabilidade na equação 3.2 NAM são calculados a partir dos dados de inteligibilidade dos monossílabos do inglês americano.

A evidência experimental para o NAM vem de experimentos de decisão lexical auditiva e de nomeação. Em ambos os casos, a densidade e frequência da vizinhança mostraram efeitos estatisticamente significativos, de acordo com as categorias extremas previstas pela equação 3.2, na direção de um maior tempo de reação para as palavras com alta densidade e alta frequência de vizinhança.

Uma adaptação possível da FWNPR na ausência de dados comportamentais sobre inteligibilidade é a utilizada por Brown (2002). Basicamente, eliminando os termos com o cálculo da probabilidade e calculando a frequência normalizada com logaritmo de base 10, temos apenas o que iremos chamar de frequência relativa (ou simplesmente **R**):

$$\mathbf{R} = \frac{\log_{10}(\mathbf{Freq}_s)}{\log_{10}(\mathbf{Freq}_s) + \sum_{j=1}^{nn} (\log_{10}(\mathbf{Freq}_{Nj}))} \quad (3.3)$$

Adotaremos para o nosso estudo parte dessa forma reduzida da equação 3.3, utilizando o termo responsável por calcular a vizinhança lexical, a saber:

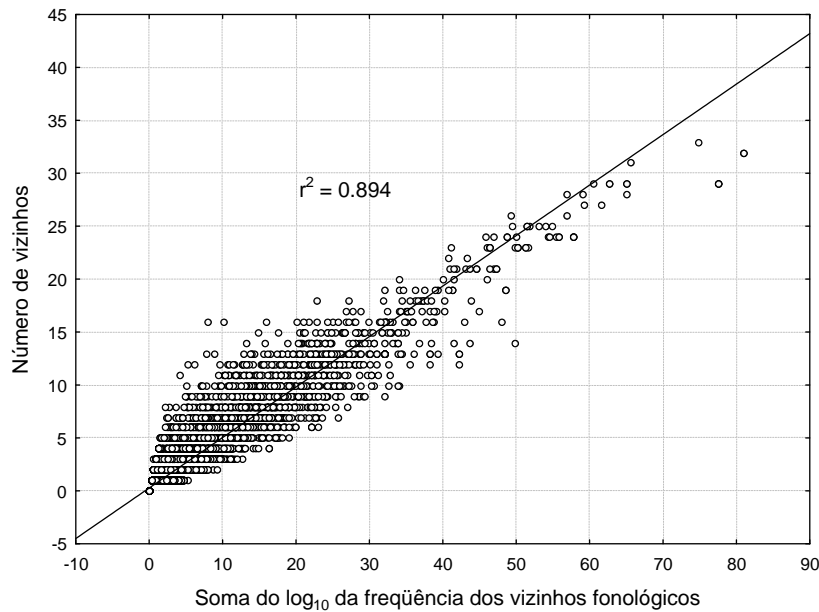


Figura 3.3: Correlação entre **SFV** e o número absoluto de vizinhos para 24578 palavras de até 14 fones do PB.

Definição 2 *Soma da frequência dos vizinhos fonológicos, abreviadamente **SFV**, na equação 3.3 representado por $\sum_{j=1}^{nn} (\log_{10}(\text{Freq}_{Nj}))$;*

Uma possível inclusão dos termos de inteligibilidade para o PB teria de sofrer uma adaptação metodológica, já que, diferentemente do inglês, há poucas palavras de conteúdo monossilábicas e a densidade e inteligibilidade tendem, portanto, a ser mais elevadas no inglês americano do que no PB.

Para o PB, a **SFV** é altamente correlacionada ($r^2 = 0.894$) com a medida mais simples de vizinhança, o número absoluto de vizinhos. Optamos por utilizar **SFV** pela possibilidade da métrica incorporar a informação de frequência dos vizinhos. É possível admitir (e comprovar) que duas palavras com o mesmo número de vizinhos, possuam vizinhos de frequências de ocorrência bastante diferentes, o que resulta em diferentes valores de **SFV**. A figura 3.3 ilustra tanto a correlação entre **SFV** e o número absoluto de vizinhos quanto a variação de **SFV** em determinado número de vizinhos para 24578 palavras com até 14 fones do PB ($t = 456$, $p < 0,001$ $r^2 = 0.9$).

Essa variação da frequência dos vizinhos fonológicos de uma palavra fica ainda mais evidente quando verificamos que o grau de correlação linear (r^2) entre a frequência de ocorrência de uma palavras e seu valor para **SFV**, calculadas excluindo-se as palavras

que não têm vizinhos, é de apenas 0,0592%³, i.e., frequência de ocorrência e vizinhança fonológica são variáveis praticamente independentes.

Uma evidência de que a informação de vizinhança parece ser relevante para o PB mostra-se na análise *a posteriori* dos dados de Françoze et alii (2002). A figura 3.4 mostra a relação entre tempo médio de reação, **SFV** e frequência de ocorrência. Podemos notar que para as palavras em que há uma maior vizinhança fonológica (medida por **SFV**) há maior predominância de um tempo de reação mais baixo que nas áreas de baixa vizinhança fonológica, especialmente para palavras de baixa frequência (canto esquerdo do gráfico). Ou seja, uma maior vizinhança lexical parece propiciar tempos de reação mais baixos mesmo em palavras de baixa frequência de ocorrência, onde se espera tempos de reação mais elevados, com exceção das palavras de baixíssima frequência, que possivelmente são interpretadas como não-palavras da

3.3 Fonotaxe probabilística

O terceiro fator que decidimos incluir como controle no segundo experimento foi a fonotaxe, isto é, os padrões combinatórios entre os segmentos (e possivelmente gestos articulatorios) da língua.

Também tradicionalmente tratada como um fenômeno do tipo categórico pela gramática gerativa, a fonotaxe também pode ser tratada em termos quantitativos. O trabalho de Coleman e Pierrehumbert (1997), por exemplo, mostra que a aceitabilidade de logatomas no inglês, medida através de julgamentos binários posteriormente convertidos em uma escala de aceitabilidade, é gradiente e está correlacionada com a informação de frequência de ocorrência das sub-partes dos logatomas. Também Frisch (1996) utiliza medidas de fonotaxe probabilística para modelar um fenômeno gramatical como o Princípio do Contorno Obrigatório no inglês e no árabe e Vitevitch e Luce (1999) revisam experimentos em que logatomas com maior probabilidade fonotática exibem um efeito facilitatório (menor TR) em tarefas psicofísicas. Em PB, o trabalho de Gama-Rossi (2002) começa a mapear as preferências fonotáticas para os falantes da língua.

Coerentemente com a inclusão dos fatores frequência (através da oposição palavra *vs.* logatoma) e vizinhança lexical, utilizaremos uma medida de fonotaxe probabilística

³A correlação de *Pearson* é, no entanto, estatisticamente significativa ($t = 31,8, p < 0,001$).

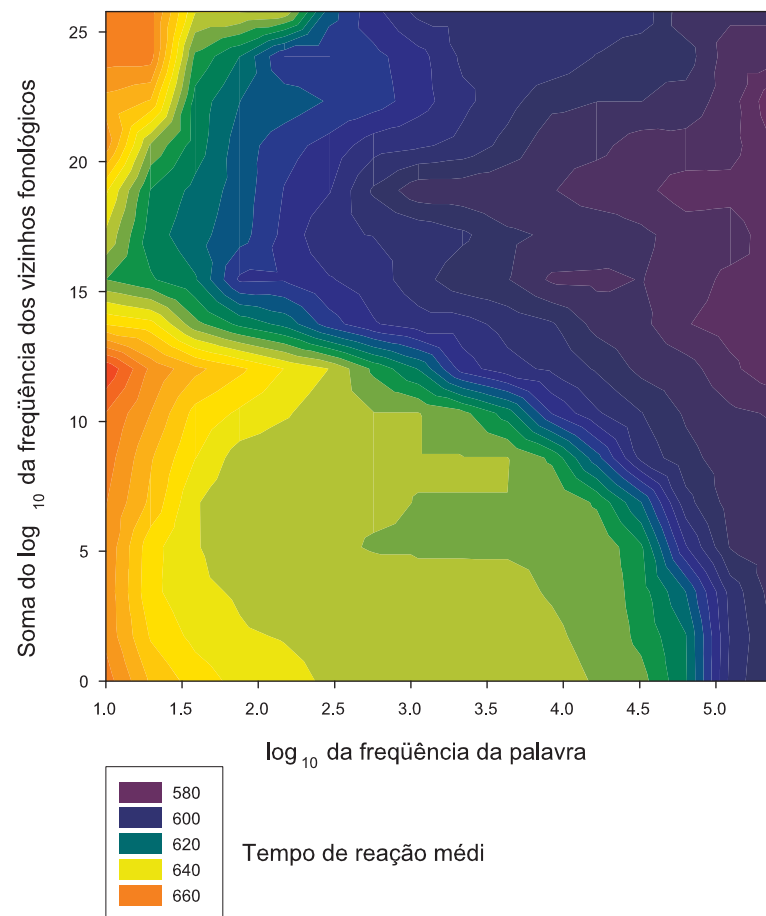


Figura 3.4: Superfície de contorno (*spline*) para frequência de ocorrência e **SFV** por tempo de reação

bastante simples, mas que utiliza informação de frequência de ocorrência. Ela é derivada das medidas utilizadas na literatura sobre fonotaxe, mais especificamente Luce e Pisoni (1998), porém reduzida a um mínimo por uma questão de simplicidade. Até o presente momento, não dispomos de medidas testadas empiricamente para o PB e não nos pareceu adequado utilizar as mesmas medidas do inglês, em que há um número de monossílabos muito maior que o português.

A métrica adotada buscou captar a intuição de que uma palavra formada por seqüências fônicas mais frequentes é fonotaticamente mais aceitável. A unidade que escolhemos para calcular as frequências das seqüências fônicas foi o difone, definido como uma seqüência contígua de dois fones. Uma palavra como *casa* ([ˈkazɐ]), por exemplo, é formada pelos difones [ka] + [az] + [zɐ].

A métrica é expressa na equação 3.4:

$$\sum_{j=1}^n \left(\log_{10} \left(\mathbf{Freq}_{diphone_j} \right) \right) \quad (3.4)$$

em que n é o número de difones da palavra, e $\mathbf{Freq}_{diphone_j}$ é a frequência de ocorrência absoluta somada de todas as ocorrências do difone no *corpus* utilizado para estimar a frequência de ocorrência dos itens lexicais no segundo experimento (seção 4.1.1).

Apesar de simples, a métrica se mostra apropriada para distinguir numericamente que uma palavra como “*chuchu*” (0.088) possui partes mais raras que uma palavra como “*para*” (0.447). Dada a sua definição, nossa métrica de fonotaxe é, assim como **SFV**, dependente do número de segmentos da palavra ou logatoma. Outra métrica de fonotaxe possível, seria o cálculo das probabilidades condicionais de transição entre bigramas (podendo ser adaptada a difones), amplamente utilizada em algoritmos de reconhecimento de fala. No entanto, não poderíamos empregar tal técnica no presente momento por não dispormos de um *corpus* de texto contínuo com uma transcrição automática satisfatória.

Capítulo 4

Experimento 2: extensão e aperfeiçoamento

O principal objetivo de um segundo experimento foi aperfeiçoar o primeiro experimento, principalmente controlando a densidade da vizinhança fonológica e aceitabilidade, aferida pela métrica de fonotaxe probabilística (seção 3.3), das palavras e logatomas. Exceto pelos resultados e discussão do experimento, as diferenças metodológicas serão explicitadas apenas quando houver alteração em relação ao primeiro experimento.

4.1 Metodologia

4.1.1 *Corpora* utilizados

Corpora para cálculo de frequências

Para este segundo experimento, além de um novo *corpus* de gravações, utilizamos outros dois *corpora* para calcular a frequência de ocorrência dos itens lexicais e difones e para o cálculo da vizinhança fonológica.

O primeiro *corpus* que utilizamos foi o *corpus* do Minidicionário Aurélio, que é a mesma base utilizada em outros levantamentos estatísticos realizados no LAFAPE (ALBANO, 2001a, 2002). Este *corpus* contém 27074 formas básicas (sem incluir flexões e muitas derivações) com uma transcrição fonológica larga e marcações de sílaba e acento tônico. Com este *corpus*, podemos precisar os vizinhos lexicais de acordo com a **Definição 1**, levando em conta inclusive a posição do acento lexical.

Para estimar a frequência de ocorrência dos itens lexicais e dos difones, utilizamos um segundo *corpus*, o CETENFolha¹ (Corpus de **E**xtractos de **T**extos **E**lectrónicos NILC/Folha de S. Paulo). O CETENFolha é um *corpus* apenas de língua escrita, contendo as 365 edições da Folha de São Paulo do ano de 1994. Possui aproximadamente 24 milhões de formas (*tokens*) e é suficientemente robusto para estimar a frequência de ocorrência das palavras no PB, ainda que não idealmente, por ser uma fonte exclusivamente de língua escrita, com predominância de estilo jornalístico e contendo dados apenas do ano de 1994. O CETENFolha contém formas lexicais em forma ortográfica, incluindo formas flexionadas e derivadas, e inclui também palavras em outras línguas, erros de digitação e nomes próprios. A remoção manual de imprecisões num *corpus* de tal magnitude é inviável, mas, pelo mesmo motivo, tais imprecisões tendem a ter um efeito estatístico desprezível. A partir do texto do CETENFolha, construímos uma lista de palavras com a respectiva frequência de ocorrência bruta de cada palavra e uma lista com a frequência somada de todas as ocorrências de cada difone.

***Corpus* de locuções**

O *corpus* de locuções coletado para o segundo experimento inclui locuções para contemplar o novo par de logatomas *vápa/vapá* e verbos *paga/pagar* (seção 4.1.3). Inclui ainda contexto adicional que permitisse que todas as vogais pós-tônicas estivessem em posição final, e.g., “Digo [ˈpape ˈgate]” (apêndice 5.3). O *corpus* também foi utilizado para extrair todas as médias de duração e formantes que serviram como parâmetro nas concatenações.

4.1.2 Sujeitos

Para o segundo experimento, aumentamos o número de participantes para 18, sendo a maioria estudantes universitários com idades entre 17 e 30 anos. Todos os participantes tomaram parte voluntariamente no experimento e assinaram um termo de participação (ver apêndice 5.3) antes do início do teste. Os dados de todos os sujeitos foram incluídos na análise estatística.

¹Disponível para download gratuitamente, necessitando apenas de autorização, em <http://acdc.linguateca.pt/cetenfolha/>.

4.1.3 Estímulos

Os pares de verbos e logatomas do segundo experimento foram selecionados de modo a estarem, tanto quanto possível, na mesma faixa de aceitabilidade (seção 4.2.2), e densidade de vizinhança (seção 3.2), ambas faixas definidas por quartis da distribuição normal. As formas verbais de “pagar” permaneceram e utilizamos o par de logatomas [‘vape] e [va‘pa] com respectivas variantes:

1. A partir da forma-base oxítônica: [va‘pa], [və‘pa] e [va‘pə];
2. A partir da forma-base paroxítônica: [‘vape], [‘vəpe] e [‘vapa].

A concatenação dos estímulos foi realizada com a técnica de síntese *PSOLA*, disponibilizada pelo programa *Praat*, versão 4 (BOERSMA; WEENINK, 1996). O motivo da mudança de técnica de síntese justifica-se apenas por uma facilidade computacional, já que a tarefa de concatenação foi bastante simplificada com a utilização do *Praat*. A qualidade da concatenação permaneceu boa e sem artefatos audíveis.

Para o segundo experimento, também aperfeiçoamos a escolha das vogais a serem utilizadas na concatenação, utilizando aqueles exemplares cujos valores dos formantes se desviassem menos da média dos valores no *corpus* coletado. A vogal selecionada foi a que apresentava o menor valor absoluto dos desvios dos formantes em relação à média, relação representada pela equação 4.1.

$$\mathbf{O}_t = \sum_{i=1}^3 \left(\left| \frac{\mathbf{f}_i - \boldsymbol{\mu}_i}{\boldsymbol{\sigma}_i} \right| \right) \quad (4.1)$$

em que \mathbf{O}_t é a soma dos desvios de todos as frequências dos formantes em relação à média, i é a ordem do formante até o terceiro, i.e., \mathbf{F}_1 , \mathbf{F}_2 e \mathbf{F}_3 , \mathbf{f}_i é o valor da frequência do formante i da amostra, $\boldsymbol{\mu}_i$ é a média do valor do formante i do *corpus* e $\boldsymbol{\sigma}_i$ é o desvio padrão do formante i do *corpus*. Optamos por não atribuir peso maior a nenhum dos formantes no cálculo de \mathbf{O}_t por ser tal atribuição um tanto arbitrária. A amostra escolhida dentre as disponíveis no *corpus* foi aquela com o menor valor de \mathbf{O}_t .

4.2 Controles

4.2.1 Controle de vizinhança fonológica

Para equilibrar a vizinhança fonológica de acordo com a métrica **SFV**, escolhemos pares de palavras e logatomas em que cada uma das formas do par mínimo ([*'page*] e [*'vape*]; [*pa'ga*] e [*va'pa*]) estivesse no mesmo quartil. Ainda que a distribuição da métrica não seja completamente normal, nos valemos da tendência à normalidade que apresentam. As figuras 4.1 e 4.2 mostram a distribuição das formas escolhidas numa distribuição normal teórica para a faixa de variação empírica para quatro e cinco segmentos.

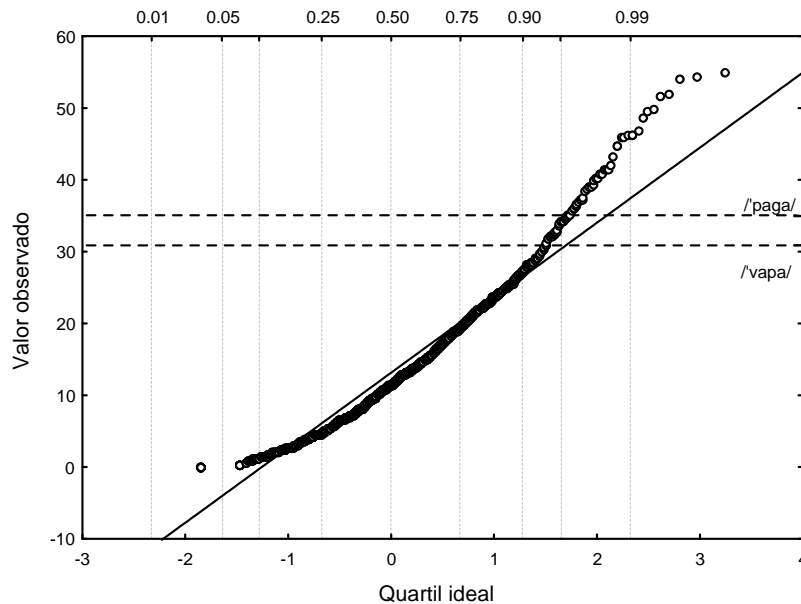


Figura 4.1: Distribuição de [*'page*] e [*'vape*] na distribuição normal teórica de **SFV** para quatro segmentos.

4.2.2 Controle de fonotaxe probabilística

O controle da métrica de fonotaxe probabilística obedeceu aos mesmos critérios adotados no controle de **SFV**. As figuras 4.3 e 4.4 ilustram a posição numérica das formas dentre os valores de variação possíveis.

4.2.3 Variáveis

Em relação ao experimento 1, alteramos apenas algumas variáveis independentes. Excluímos os dois níveis distintos de duração, que não apresentaram diferenças estatisticamente

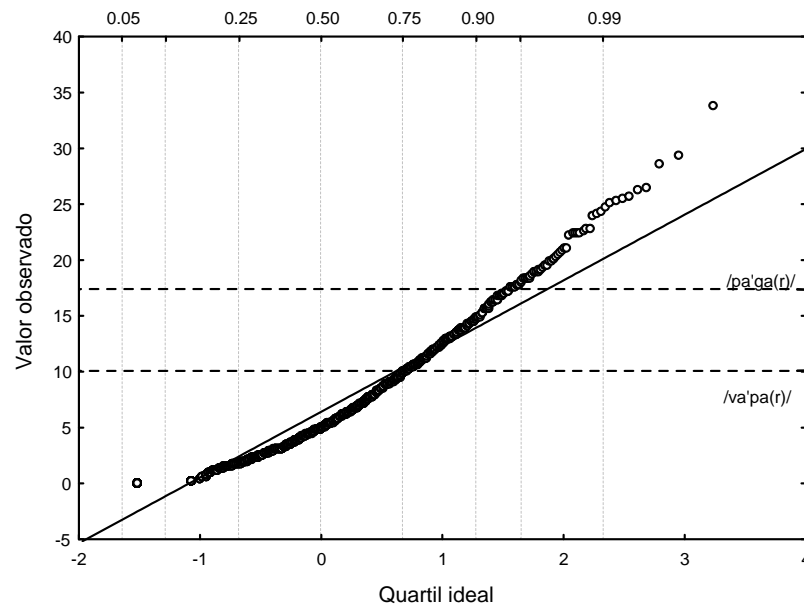


Figura 4.2: Distribuição de [pa'ga] e [va'pa] na distribuição normal teórica de **SFV** para cinco segmentos.

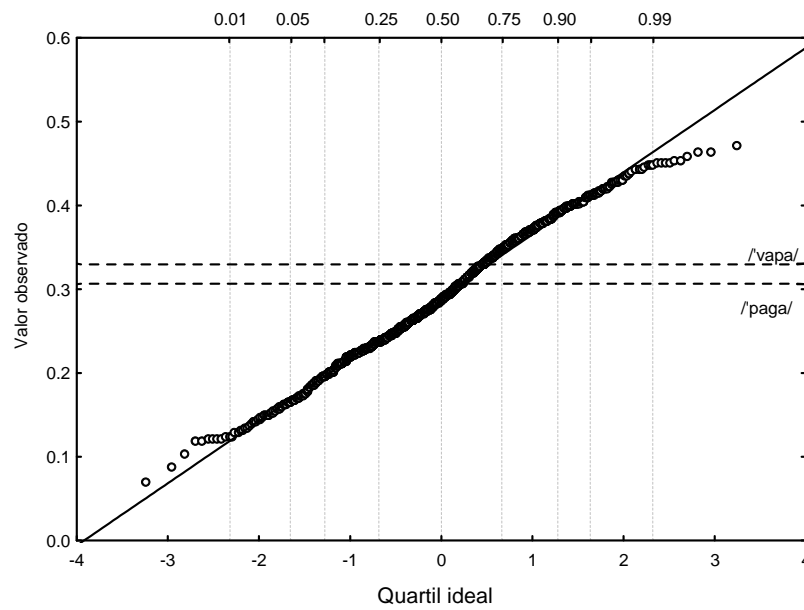


Figura 4.3: Distribuição de ['pagə] e ['vapə] na distribuição normal teórica de Fonotaxe probabilística para quatro segmentos.

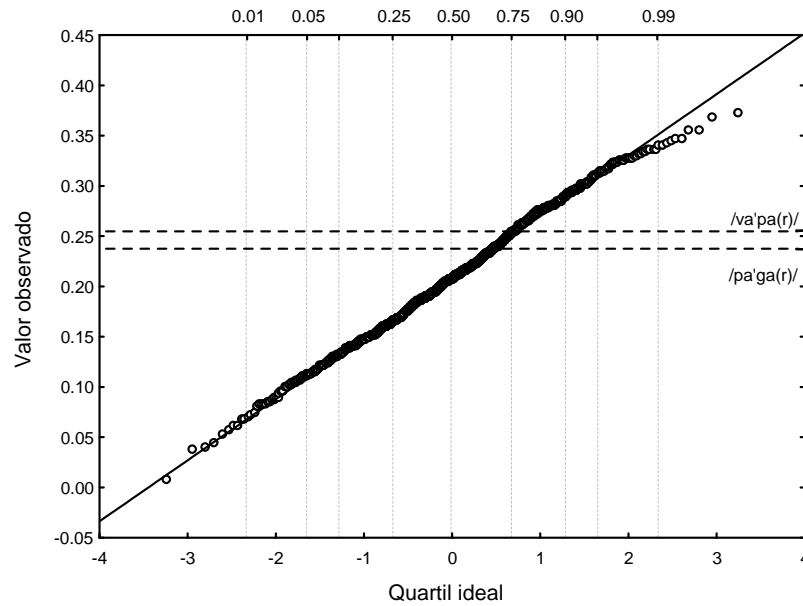


Figura 4.4: Distribuição de [pa'ga] e [va'pa] na distribuição normal teórica de Fonotaxe probabilística para cinco segmentos.

significativas, e incluímos dois níveis de amplitude: a) vogais com amplitude relativa canônica; e b) vogais com amplitude relativa iguais (ver 4.1.3). As vogais da condição *b* foram obtidas igualando a amplitude máxima da forma de onda no mesmo nível da vogal mais forte do estímulo. Já as vogais da condição *a* foram ajustadas para apresentarem uma diferença dos picos de amplitude na média das vogais do *corpus* em posição similar.

4.2.4 Procedimentos

No segundo experimento, alteramos o intervalo entre os estímulos nas tríades para 500 ms para evitar que os sujeitos possivelmente tendessem a responder as tríades como seqüências puramente fonéticas (WHALEN, D., 2002). O número de repetições de cada agrupamento de tríades foi reduzido para oito, totalizando 192 formas a serem respondidas por sujeito em cada sessão.

4.3 Análise Estatística

A análise estatística do segundo experimento seguiu as mesmas linhas do experimento 1, apenas com uma modificação no fator de normalização.

O fator λ foi ajustado em 0,1932. A tabela 4.1 apresenta a estatística descritiva básica dos dados brutos e normalizados e a tabela 4.2 apresenta o padrão de erros para os sujeitos e blocos experimentais (seção 2.2.1).

	média	mediana	σ	faixa	assimetria	curtose
TR	631	563	314	1967	1,31	2,2
TR $\lambda = 0,1932$	12,5	12,4	1,6	12,3	0,06	0,44

Tabela 4.1: Valores de média, mediana, desvio padrão, faixa de variação (em ms), assimetria e curtose para os tempos de reação brutos e normalizados, experimento 2.

	média	σ	mínimo	máximo	erro geral %
Blocos (8)	3,5	1,4	0	10	15%
Sujeitos (18)	28,4	10,6	10	50	15%

Tabela 4.2: Padrão de erros para os blocos e sujeitos, experimento 2.

4.4 Resultados

4.4.1 Resultados de tempo de reação

A ANOVA de medidas repetidas para os dados de TR apresentou significância apenas para o fator acento lexical ($F(1,17) = 12.39, p = 0.002$). Novamente, os sujeitos discriminaram mais rapidamente alofones contidos em formas com acento paroxítono do que aqueles contidos em acento oxítono. Os demais fatores e interações entre fatores não apresentaram diferença estatística significativa para TR.

4.4.2 Resultados de padrão de acertos/erros

Para o padrão de acertos/erros, o fator lexicalidade ($\chi^2(1) = 7,99, p = 0,0047$) e o fator acento lexical ($\chi^2(1) = 48,28, p < 0,001$) apresentaram diferenças estatisticamente significativas. Os sujeitos discriminaram com maior precisão os alofones contidos em palavras e em formas paroxítonas. A tabela 4.3 sumariza o padrão de respostas corretas e incorretas para os fatores independentes binários.

Nas análises dois a dois das tríades, obtivemos os seguintes resultados:

1. Tríades com uma forma incorreta \times tríades com duas formas incorretas: ($\chi^2(1) = 99,80, p < 0,001$);

	acertos	erros
palavras	1502	226
logatomas	1443	285
paroxítonos	1545	183
oxítonos	1400	382
amplitude constante	1480	248
amplitude ajustada	1465	263

Tabela 4.3: Padrão de acertos/erros para os fatores independentes binários, experimento 2.

2. Tríades com uma forma incorreta \times tríades com três formas incorretas: ($\chi^2(1) = 162,19, p < 0,001$);
3. Tríades com duas formas incorretas \times tríades com três formas incorretas: ($\chi^2(1) = 9,57, p = 0,002$);

Em todas as comparações, novamente obtivemos diferenças significativas no número de acertos e erros a depender da quantidade de estímulos incorretos presentes nas tríades. Há maior número de erros quanto menor o número de formas incorretas. A figura 4.5 exibe graficamente os resultados dos testes estatísticos.

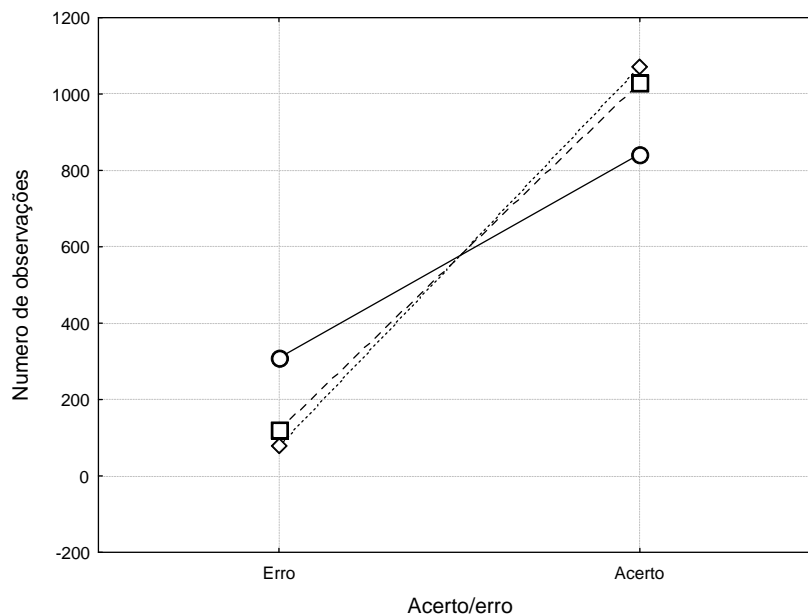


Figura 4.5: Quantidade de acertos/erros para as tríades classificadas por quantidade de formas incorretas, experimento 2.

4.5 Discussão

Os resultados do segundo experimento, realizado com um controle de vizinhança fonológica e fonotaxe e com um maior número de sujeitos, indicam que a percepção de alofones não parece ser indiferente ao contexto. Resumidamente, os sujeitos discriminaram mais rapidamente e foram mais precisos ao discriminar alofones em formas paroxítonas. Os sujeitos foram mais precisos ao discriminar os alofones quando as tríades apresentavam maior número de formas incorretas e também foram mais precisos ao discriminar alofones contidos em palavras da língua. Este último, um resultado oposto ao encontrado no experimento 1 (seção 2.3).

O resultado relativo ao tempo de reação pode a princípio parecer previsível: os sujeitos discriminam mais rapidamente formas paroxítonas porque podem decidir prematuramente com qual forma associar a terceira ao ouvir a primeira sílaba do último estímulo. Nas oxítonas, os sujeitos precisam ouvir a última sílaba para tomar a decisão. Apesar de possível, essa interpretação é desfavorecida quando cruzamos os dados de acertos/erros, que apontam uma maior dificuldade em diferenciar alofones em formas oxítonas (com um resultado marginal no experimento 1 e na mesma direção), lembrando que os dados são coletados simultaneamente. Não só os sujeitos discriminam mais rapidamente alofones em formas de acento canônico, como o fazem mais precisamente num nível acima do aleatório. Este resultado, fosse um simples artefato da apresentação, poderia ser indiferente à acuidade da resposta, como o fator duração no experimento 1 e com o fator amplitude no segundo experimento o são e sem apresentar um correlato na medida de TR.

Já os resultados do padrão de erros para o fator lexicalidade não permitem maiores conclusões, dado o conflito de resultados entre os experimentos. Talvez o que estejamos rotulando por lexicalidade (ou extremos de frequência) ainda esteja encobrindo outras variáveis. O que seria até esperado, já que fatores como fonotaxe e vizinhança lexical são características que ajudam a definir o quanto uma forma se assemelha mais ou menos a outras palavras da língua. Lexicalidade, nessa perspectiva, dificilmente poderia ser considerada uma variável categórica.

Capítulo 5

Discussão

5.1 Discussão dos resultados

Retomando a nossa pergunta inicial, analisemos em que medida os resultados encontrados favorecem uma especificação lexical da alofonia discreta. Mais especificamente: quais fatores dentre os investigados diferenciam a percepção dos alofones e em que direções.

O fator lexicalidade, que no experimento de (WHALEN, D. H.; BEST; IRWIN, 1997) foi o que apresentou diferença significativa para a percepção do grau de acento estrangeiro (seção 1.3) não separa os tempos de reação na discriminação dos alofones. Se no primeiro experimento (capítulo 2) essa ausência de diferença poderia ser atribuída a uma falta de controle experimental de vizinhança fonológica e tipicidade, no segundo experimento (capítulo 4), tais parâmetros foram controlados com rigor. Ainda assim, os resultados negativos quanto à lexicalidade permaneceram. Como os fatores frequência de ocorrência, fonotaxe e vizinhança fonológica influenciam a percepção de fala, tendo o experimento 2 controlado os dois últimos fatores minimamente, a única variável manipulada que poderia afetar o processo perceptual era a frequência de ocorrência, especificada de forma categórica na distinção palavra \times logatoma.

Tomado isoladamente, este resultado poderia apontar na direção da ausência de influência do fator frequência dos itens lexicais na percepção da alofonia. No entanto, há uma importante diferença em nossos experimentos em relação aos efeitos de frequência na percepção de fala: a complexidade da tarefa. Num experimento típico de decisão lexical auditiva, há uma forma sem alterações fonéticas a ser discriminada. Nos nossos experimentos, a discriminação envolvia no mínimo uma forma concatenada com alofones em

posição incorreta. É plausível supor, portanto que, em nossos experimentos, o acesso lexical não seria tão imediato quanto na tarefa de decisão lexical simples por envolver sempre formas que não pertencem estritamente à língua. Na parte da tarefa menos sensível a um reconhecimento imediato, o padrão de acertos/erros, há o resultado que indica que formas baseadas em uma forma mais freqüente tornam mais apurada a distinção entre os alofones [a, ɐ]. Na mesma direção, temos o resultado da distinção entre as tríades com com uma, duas ou três formas com alofones em posição incorreta, que mostra que a discriminação melhora com um menor número de formas com alofones incorretos, o que sugere uma gradiência do fenômeno na direção da melhora do desempenho diretamente proporcional ao aumento do número de formas corretas a serem discriminadas.

Ainda em relação à distinção de lexicalidade, é necessário lembrar que nossa medida de fonotaxe probabilística é bastante simplificada (seção 3.3, possivelmente não capturando da melhor forma possível as diferenças entre formas preferidas e repelidas pela língua. Estudos como os de (FRISCH; LARGE; PISONI, 2000), (VITEVITCH; LUCE, 1999) e (SAFFRAN; ASLIN; NEWPORT, 1996) mostram uma facilitação do acesso lexical com o aumento da tipicidade fônica de palavras e de logatomas. Estes estudos mostram que a fronteira entre uma palavra da língua, um logatoma com seqüências bastante típicas e logatomas com seqüência que correm com baixa nenhuma freqüência não é claramente delimitada. Qual a métrica mais apropriada a capturar ainda é uma questão em aberto na literatura. Bailey e Hahn (2001), por exemplo, testaram seis métricas distintas, incluindo bigramas, trigramas e métricas silábicas e verificaram que todas as métricas têm uma capacidade limitada ($r^2 = 0.19$ como máximo) de explicar a variação de dados de julgamentos de lexicalidade para o inglês. Métricas baseadas em gestos articulatórios podem se mostrar um teste futuro importante para a FAR.

Para o fator acento lexical, os resultados de ambos os experimentos para TR apontam uma diferença significativa entre os dois níveis estudados, com um menor TR para as formas paroxítonas. O segundo experimento confirma esse resultado na mesma direção com um menor numero de erros, i.e., uma maior facilidade na tarefa, também para formas paroxítonas. A influência do padrão canônico de acentuação da língua é evidência adicional de que a percepção da alofonia não se dá de maneira uniforme na língua. Formas acentuais canônicas (mais freqüentes) na língua favorecem a discriminação da alofonia de maneira robusta, sugerindo novamente que as diferenças alofônicas podem estar sensíveis

à percepção. Surge, portanto, a necessidade de um léxico em que algum detalhe fonético fino possa ser previsto.

5.2 Fonologia Articulatória e teoria de exemplares

Com os resultados de ambos os experimentos, temos, além da possibilidade dos alofones estarem especificados lexicalmente, a necessidade de pensar em uma extensão à FAR que incorpore variáveis distribucionais como frequência de ocorrência, vizinhança e fonotaxe probabilística.

Um modelo fonológico que propõe considerar tanto o detalhe fonético quanto a informação distribucional é proposto por Pierrehumbert ([PIERREHUMBERT, J., 2002, 2001](#); [PIERREHUMBERT, J. B., 2003](#)). A proposta de Pierrehumbert baseia-se nas teorias de exemplares de percepção estendidas à produção de fala. Na teoria de exemplares, as categorias fônicas são representadas na memória de longo termo pela acumulação de exemplares (*tokens*) de cada categoria, sendo que cada categoria pode variar nas várias dimensões pertinentes à sua especificação (e.g., F_1 , F_2 , duração). Categorias com mais exemplares teriam, então, acesso lexical mais imediato que categorias com poucos exemplares na memória de longo termo.

Pierrehumbert ([2001](#)) realizou simulações computacionais de mudanças fonológicas que a redução vocálica a *schwa* e o apagamento de /t/, que mostraram como uma mudança sistemática num parâmetro fonético produzem com o passar do tempo categorias novas. Estas categorias novas, a princípio emergem inicialmente em palavras de alta frequência, como consequência destas possuírem mais exemplares produzidos e percebidos, de modo similar ao que é atestado empiricamente para o inglês ([BYBEE, 2000](#); [LABOV, 1981](#)).

Apesar de incluir com sucesso a informação sobre o detalhe lexical e como esse detalhe lexical pode diferenciar o léxico de maneira não-uniforme, a proposta de Pierrehumbert ainda não desenvolveu uma especificação fonética tão poderosa quanto a da FAR, limitando-se à uma representação segmental da matéria fônica. Uma solução possível para uma possível integração de ambas teorias é considerar também os gestos articulatórios como dimensões relevantes na especificação lexical, de maneira semelhante à proposta de Albano ([2001](#)).

Nossos resultados, particularmente o menor TR e menor taxa de erros para formas paroxítonas, são compatíveis com a possibilidade de incorporação da proposta de Pierrehumbert à FAR. Formas com acento lexical canônico, as que apresentam mais exemplares na memória de longo termo, sugerem uma segmentação do léxico, na medida em que diferenciam a percepção da alofonia e introduzem uma possível nova dimensão na especificação lexical. Fatores como frequência de ocorrência, fonotaxe probabilística e vizinhança fonológica também poderiam ser consideradas dimensões adicionais na representação lexical. Ainda que um léxico mais detalhado seja contrário à parcimônia dos modelos gerativos e menos desejável em termos de simplicidade científica, ele parece se fazer necessário para prever alguns fatos empíricos, como os apresentados e discutidos neste trabalho, que desafiam a visão de uniformidade do léxico.

5.3 Lexicalidade: quais as fronteiras?

Fica claro a partir dos nossos resultados e dos trabalhos resenhados que o que rotineiramente se entende por lexicalidade não parece ser um atributo binário das formas fônicas (gestuais) da língua e que provavelmente é determinado por mais de uma dimensão. O modelo lexical que vislumbramos é o de um léxico em que detalhes fonéticos específicos convivam com processos fônicos produtivos, ora favorecendo-os e ora bloqueando-os. O desafio que se abre é o de investigar o que define as categorias fônicas, e qual é o detalhamento plausível, i.e. quais as dimensões constitutivas dessas formas.

Os experimentos aqui reportados podem ser ampliados para incluírem uma variação de vizinhança fonológica, fonotaxe probabilística e possivelmente outros fatores, para testar até que ponto tais fatores podem esclarecer o nível de detalhe fonético (necessário) a ser especificado nas entradas lexicais.

A redução vocálica mostrou-se um fenômeno particularmente apropriado a discutir a possibilidade de especificação dos alofones na entrada lexical por ser, ao mesmo tempo, um fenômeno com manifestação gradiente e categórica (ALBANO, 2001b). A manifestação gradiente podendo ser explicada a partir dos princípios 1, variação do grau de sobreposição, e 2, variação da magnitude espacial e temporal, e a manifestação categórica demandando a especificação pelo contraste de gestos articulatórios. Essa convivência de ambos os princípios atuando nos mesmos tipos de processos fônicos é que permite à FAR

ser uma proposta de teoria gramatical mais completa e abrangente e não simplesmente de um modelo de implementação fonética desvinculado de outros aspectos gramaticais.

Referências bibliográficas

- ALBANO, E. C. et al. A interface fonética-fonologia e a interação prosódia-segmentos. In: ESTUDOS Lingüísticos XXVII: Anais do Grupo de Estudos Lingüísticos do Estado de São Paulo (GEL). São José do Rio Preto: UNESP-IBILCE, 1998. p. 135–143.
- ALBANO, E. C. **O gesto e suas bordas: esboço de Fonologia acústico-articulatória do português brasileiro**. Campinas: Mercado de Letras, 2001.
- O gramatical e o abstrato num modelo fonético/fonológico dinâmico. In: — . **Anais do II Congresso Internacional da ABRALIN**. Fortaleza: [s.n.], 2001.
- **V-to-’V phonotactics, syllable structure, and morphological productivity**. Yale University, New Haven: [s.n.], jun. 2002. Poster apresentado no *Labphon 8*.
- ALBANO, E. C. Criatividade e gradiência num léxico sem derivações. In: CABRAL, L. G.; MORAIS, J. (Ed.). **Investigando a linguagem: ensaios em homenagem a Leonor Scliar-Cabral**. Florianópolis: Editora Mulheres, 1999. p. 35–54.
- AQUINO, P. A. **O papel das vogais reduzidas pós-tônicas na construção de um sistema de síntese concatenativa para o português do Brasil**. 1997. Diss. (Mestrado) – Instituto de Estudos da Linguagem — Universidade Estadual de Campinas, Campinas.
- ARAGÃO, M. As pesquisas dialetais no Brasil: aspectos fônicos – as vogais. In: ANAIS do II Congresso Nacional da ABRALIN. [S.l.: s.n.], 1999. Disponível em http://sw.npd.ufc.br/abralin/anais_con2nac_sum.html.
- BAILEY, T. M.; HAHN, U. Determinants of wordlikeness: phonotactics or lexical neighborhoods? **Journal of Memory and Language**, v. 44, p. 586–591, 2001.

- BOERSMA, P.; WEENINK, D. Praat, a system for doing phonetics by computer. **Technical Report, Institute of Phonetic Sciences of the University of Amsterdam**, v. 132, n.a. 1996.
- BOX, G. E. P.; COX, D. R. An Analysis of Transformations. **Journal of the Royal Statistical Society**, B-26, p. 211–234, 1964.
- BROWMAN, C. P.; GOLDSTEIN, L. Articulatory phonology: an overview. **Phonetica**, v. 49, p. 155–180, 1992.
- BROWN, R. A. Effects of lexical confusability on the production of coarticulation. **UCLA Working Papers in Phonetics**, v. 100, p. 73–106, 2002.
- BYBEE, J. The phonology of the lexicon: evidence from lexical diffusion. In: BARLOW, M.; KEMMERER, S. (Ed.). **Usage-Based Models of language**. Stanford: CSLI, 2000.
- CARPENTER, R. H. S.; WILLIAMS, M. L. L. Neural computation of log likelihood in the control of saccadic eye movements. **Nature**, v. 377, p. 59–62, 1995.
- CHOMSKY, N.; HALE, M. **The sound pattern of English**. Nova Iorque: Harper & Row, 1968.
- COLEMAN, J.; PIERREHUMBERT, J. Stochastic Phonological Grammars and Acceptability. In: — . **Proceedings of the 3rd Meeting of the ACL Special Interest Group in Computational Phonology**. Somerset, NJ: [s.n.], 1997. p. 49–56.
- COLTHEART, M. Access to the internal lexicon. In: **ATTENTION & Performance VI**. New York: Academic Press, 1977. p. 535–555.
- D'ANGELIS, W. d. R. **Traços de modo e modos de traçar geometrias: línguas Macro-Jê e teoria fonológica**. 1998. Tese (Doutorado) – Universidade Estadual de Campinas, Campinas, SP.
- DUPOUX, E. et al. Epenthetic vowels in Japanese: a perceptual illusion? **Journal of Experimental Psychology: Human Perception and Performance**, v. 25, p. 1568–1578, 1999.
- FORSTER, K. I.; FORSTER, J. C. DMDX: A Windows display program with millisecond accuracy. **Behavior Research Methods, Instruments, & Computers**, v. 35, p. 116–124, 2003.

- FOWLER, C. A. An event approach to the study of speech perception from a direct-realist perspective. **Journal of Phonetics**, v. 14, p. 3–28, 1986.
- Coarticulation and theories of extrinsic timing. **Journal of Phonetics**, v. 8, p. 113–133, 1980.
- FRANÇOZO, E. et al. **Conhecimento lexical e processamento de linguagem**. [S.l.: s.n.], 2002. Submetido aos Anais do 50o GEL.
- FRISCH, S. **Similarity and Frequency in Phonology**. 1996. Tese (Doutorado) – Northwestern University, Evanston, IL. Disponível em: citeseer.nj.nec.com/frisch96similarity.html.
- FRISCH, S.; LARGE, N. R.; PISONI, D. B. Perception of wordlikeness: effects of segment probability and length on the processing of nonwords. **Journal of Memory and Language**, v. 42, p. 481–496, 2000.
- GAMA ROSSI, A. Julgamento de gramaticalidade da fonotaxe do português: dados de um experimento. **Intercâmbio**, v. 11, p. 157–164, 2002.
- GLASS, G. V.; PECKHAM, P. D.; SANDERS, J. R. Consequences of failure to meet the assumptions underlying the fixed effects analysis of variance and covariance. **Review of Educational Research**, v. 42, p. 237–288, 1972.
- HOWES, D. H.; SOLOMAN, R. L. Visual duration threshold as a function of word probability. **Journal of Experimental Psychology**, v. 41, p. 401–410, 1951.
- HUFFMAN, M. K. Phonetic variation in intervocalic onset /l/’s in English. **Journal of Phonetics**, v. 25, p. 115–141, 1997.
- KELSO, J. A. S. **Dynamic Patterns: The Self-Organization of Brain and Behavior**. Cambridge, MA: MIT Press, 1995.
- KENSTOWICZ, M. **Phonology in Generative Grammar**. Cambridge, MA: Blackwell Publishers, 1994.
- Uniform exponence: exemplification and extension. **ROA-218**, 1998. Rutgers Optimality Archive, <http://roa.rutgers.edu/>.
- KOCHETOV, A. Syllable position effects and gestural organization: articulatory evidence from russian. Paper presented at the 8th Conference on Laboratory Phonology. [S.l.], 2002.

KUCERA, F.; FRANCIS, W. **Computational analysis of present day American English**. Providence, RI: Brown University Press, 1967.

LABOV, W. Resolving the Neogrammarian controversy. **Language**, v. 57, p. 267–309, 1981.

LIBERMAN, A. M.; MATTINGLY, I. G. The motor theory of speech perception revised. **Cognition**, v. 21, p. 1–36, 1985.

LÖFQVIST, A. Theories and models of speech production. In: HARDCASTLE, W.; LAVER, J. (Ed.). **The Handbook of Phonetic Sciences**. Oxford: Blackwell, 1997. p. 404–426.

LUCE, P. A.; CHARLES-LUCE, J.; MCLENNAN, C. Representational specificity of lexical form in the production and perception of spoken words. In: PROCEEDINGS of the XIVth International Congress of Phonetic Sciences. São Francisco: [s.n.], 1999. v. 1, p. 1889–1892.

LUCE, P. A.; PISONI, D. B. Recognizing spoken words: the neighborhood activation model. **Ear & Hearing**, v. 19, p. 1–36, 1998.

MATTOSO CÂMARA JR., J. **Estrutura da Língua Portuguesa**. Petrópolis: Vozes, 1970.

MAX, L.; ONGHENA, P. Some issues in the statistical analysis of completely randomized and repeated measures designs for speech. **Journal of Speech, Language, and Hearing Research**, v. 42, p. 261–270, 1999.

NORRIS, D. G.; MCQUEEN, J. M.; CUTLER, A. Merging information in speech recognition: Feedback is never necessary. **Behavioral and Brain Sciences**, v. 23, p. 299–325, 2000.

PIERREHUMBERT, J. Exemplar dynamics: Word frequency, lenition, and contrast. In: BYBEE, J. L.; HOPPER, P. (Ed.). **Frequency and the Emergence of Language Structure**. Amsterdam: John Benjamins, 2001. p. 137–157.

— Word-specific phonetics. In: GUSSENHOVEN, C.; WARNER, N. (Ed.). **Laboratory Phonology VII**. Berlin: Mouton de Gruyter, 2002. p. 137–157.

- PIERREHUMBERT, J. B. Probabilistic Phonology: Discrimination and Robustness. In: BOD, R.; HAY, J.; JANNEDY, S. (Ed.). **Probability Theory in Linguistics**. Cambridge, MA: MIT Press, 2003. n.d.
- PORT, R. F.; GELDER, T. van (Ed.). **Mind as motion: explorations in the dynamics of cognition**. Cambridge, MA: The MIT Press, 1995.
- SAFFRAN, J. R.; ASLIN, R. N.; NEWPORT, E. L. Statistical learning by 8-month-old infants. **Science**, v. 274, p. 1926–1928, 1996.
- SAMUEL, A. G. Lexical uniqueness effects on phonemic restoration. **Journal of Memory and Language**, v. 26, p. 36–56, 1987.
- SILVA, A. H. P. Proposta de representação dinâmica para os róticos iniciais do PB. Artigo submetido aos Anais do 50º Congresso do Grupo de Estudos Lingüísticos do Estado de São Paulo. [S.l.], 2002.
- SPROAT, R.; FUJIMURA, O. Allophonic variation in English /l/ and its implications for phonetic implementation. **Journal of Phonetics**, v. 21, p. 291–311, 1993.
- STEVENS, K. N. On the quantal nature of speech. **Journal of Phonetics**, v. 17, p. 3–46, 1989.
- TRUBETZKOY, N. S. **Principles of Phonology**. Berkeley: University of California Press, 1969.
- VAUX, B. **Flapping in English**. Cambridge, Harvard University: [s.n.], fev. 2002. Manuscrito disponível em <http://icg.harvard.edu/~sa34/nanjo/flapping.pdf>.
- VIOLARO, F.; BÖEFFARD, O. A Hybrid Model for text-to-speech synthesis. **IEEE Transactions on Speech and Audio Processing**, v. 6, p. 426–434, 1998.
- VITEVITCH, M. S.; LUCE, P. A. Probabilistics phonotactics and neighborhood activation in spoken word recognition. **Journal of Memory and Language**, p. 374–408, 1999.
- WHALEN, D. **Comunicação pessoal**. [S.l.: s.n.], jun. 2002.
- WHALEN, D. H.; BEST, C. T.; IRWIN, J. Lexical effects in the perception and production of American English /p/ allophones. **Journal of Phonetics**, v. 25, p. 501–528, 1997.

WHALEY, C. P. Word-nonword classification time. **Journal of Verbal Learning and Verbal Behavior**, v. 17, p. 143–154, 1978.

ZIPF, G. K. **Human behaviour and the principle of least-effort**. Cambridge, MA: Addison-Wesley, 1949.

Apêndices

Ficha de participação

Nome:

Idade:

e-mail (opcional):

Possui algum problema auditivo?

Meu nome é Leonardo Oliveira e sou estudante de mestrado em Lingüística.

Peço sua colaboração voluntária no experimento lingüístico que estou conduzindo para minha dissertação de mestrado com o objetivo de estudar a percepção de palavras no português do Brasil. Não prevemos qualquer tipo de risco ou desconforto na realização do experimento.

Os dados coletados serão analisados estatisticamente e serão reportados na minha dissertação de mestrado, trabalhos apresentados a congressos ou publicados em revistas científicas. Toda informação de identificação será mantida confidencialmente.

Você poderá desistir do experimento a qualquer momento se assim desejar. Após o experimento, estou a sua disposição para discutir os objetivos do experimento.

Entendo que minha participação no experimento é voluntária e que posso desistir de realizar o experimento a qualquer momento.

Campinas, de fevereiro de 2003

Assinatura participante

Assinatura experimentador

Instruções para os experimentos

Leia com atenção as instruções a seguir.

Você irá ouvir três palavras com um curto intervalo entre elas. A primeira e a segunda palavra são diferentes entre si e a terceira palavra é igual à primeira ou é igual à segunda. Você deverá responder se essa terceira palavra é igual à primeira ou à segunda utilizando os botões da parte inferior do joystick. O botão da esquerda corresponde à primeira palavra e o botão da direita corresponde à segunda palavra. A resposta deve ser dada no menor tempo possível e somente após ouvir a terceira palavra inteira. Resposta dadas

antes da apresentação completa das palavras ou depois de um intervalo muito longo não serão registradas.

Caso responda antes do término da apresentação das palavras, repita a resposta normalmente no momento certo. Caso demore a responder e apareça na tela do monitor a mensagem “Tempo esgotado”, não responda e continue o teste normalmente com a próxima seqüência de palavras. Após cada resposta aparecerá no monitor a sua resposta, isto é, se você identificou a terceira palavra como a primeira ou como a segunda. Tente responder a todas as seqüências mesmo que não tenha certeza e o mais rapidamente possível. Evite ao máximo que o tempo para resposta se esgote.

O teste está dividido em oito partes. Ao término de cada parte você poderá fazer uma pausa caso julgue necessário e ao final da quarta parte faça uma pausa um pouco maior para descansar. Mantenha a sua concentração no que está ouvindo durante todo o teste. O teste dura aproximadamente vinte minutos. Antes do início do teste haverá um treino. Em caso de dúvida, pergunte ao experimentador antes de começar o teste.

Locuções gravadas para o experimento 1

1. Diga pága.
2. Diga pápaga.
3. Diga pagá.
4. Diga dápa.
5. Diga pádaba.
6. Diga dabá.

Locuções gravadas para o experimento 2

1. Diga pága.
2. Diga pagá.
3. Diga pápa gata
4. Diga vápa.
5. Diga vapá.
6. Diga páva pata.